

Dissertation zur Erlangung des Doktorgrades
der Fakultät für Chemie und Pharmazie
der Ludwig-Maximilians-Universität München

**Massenspektrometrie-basierte Untersuchungen
nichtkanonischer DNA- und RNA-Nukleoside**

—

**Verteilung in Organen und Geweben, altersabhängige Dynamiken
und Beteiligung an der aktiven DNA-Demethylierung**

Mirko Bastian Wagner
aus
Hamburg

2016

Erklärung

Diese Dissertation wurde im Sinne von § 7 der Promotionsordnung vom 28. November 2011 von Herrn Prof. Dr. Thomas Carell betreut.

Eidesstattliche Versicherung

Diese Dissertation wurde eigenständig und ohne unerlaubte Hilfe erarbeitet.

München, den 10.05.2016

(Mirko Wagner)

Dissertation eingereicht am	10.05.2016
1. Gutachter:	Prof. Dr. Thomas Carell
2. Gutachter:	PD Dr. Stylianos Michalakis
Mündliche Prüfung am	21.06.2016

Meinen Eltern gewidmet.

Im Rahmen meiner Promotion war ich an der Entstehung der folgenden Publikationen beteiligt:

15. **M. Wagner**[†], J. Steinbacher[†], T. F. J. Kraus[†] (**gemeinsame Koautorenschaft**), S. Michalakis, B. Hackner, T. Pfaffeneder, A. Perera, M. Müller, H. A. Kretzschmar, T. Carell, *Angew. Chem. Int. Ed.* **2015**, *54*, 12511-12514. "Age-Dependent Levels of 5-Methyl-, 5-Hydroxymethyl-, and 5-Formylcytosine in Human and Mouse Brain Tissues".
14. A. Perera,[†] D. Eisen,[†] **M. Wagner**[†] (**gemeinsame Koautorenschaft**), S. K. Laube, A. F. Künzel, S. Koch, J. Steinbacher, E. Schulze, V. Splith, N. Mittermeier, M. Müller, M. Biel, T. Carell, S. Michalakis, *Cell Rep.* **2015**, *11*, 283-294. "Tet3 is Recruited by Rest for Context-Specific Hydroxymethylation and Induction of Gene Expression".
13. P. Farinelli, A. Perera, B. Arango-Gonzalez, D. Trifunovic, **M. Wagner**, T. Carell, M. Biel, E. Zrenner, S. Michalakis, F. Paquet-Durand, P. A. R. Ekström, *Cell Death Dis.* **2014**, *5*, e1558. "DNA Methylation and Differential Gene Regulation in Photoreceptor Cell Death".
12. T. Pfaffeneder,[†] F. Spada,[†] **M. Wagner**[†] (**gemeinsame Koautorenschaft**), C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schürmann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schär, M. Müller, T. Carell, *Nat. Chem. Biol.* **2014**, *10*, 574-581. "Tet Oxidizes Thymine to 5-Hydroxymethyluracil in Mouse Embryonic Stem Cell DNA".
11. C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. T. C. Jansen, C. Bauer, M. Münzel, **M. Wagner**, M. Müller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, *Cell* **2013**, *152*, 1146-1159. "Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives".
10. C. Brandmayr,[†] **M. Wagner**[†], T. Brückl,[†] (**gemeinsame Koautorenschaft**) D. Globisch, D. Pearson, A. C. Kneuttinger, V. Reiter, A. Hienzsch, S. Koch, I. Thoma, P. Thumbs, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 11162-11165. "Isotope-Based Analysis of Modified tRNA Nucleosides Correlates Modification Density with Translational Efficiency".
9. T. Carell, C. Brandmayr, A. Hienzsch, M. Müller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, **M. Wagner**, *Angew. Chem. Int. Ed.* **2012**, *51*, 7110-7131. "Structure and Function of Noncanonical Nucleobases".
8. V. Reiter, D. M. S. Matschkal, **M. Wagner**, D. Globisch, A. C. Kneuttinger, M. Müller, T. Carell, *Nucleic Acids Res.* **2012**, *40*, 6235-6240. "The CDK5 repressor CDK5RAP1 is a methylthiotransferase acting on nuclear and mitochondrial RNA".
7. T. F. J. Kraus, D. Globisch, **M. Wagner**, S. Eigenbrod, D. Widmann, M. Münzel, M. Müller, T. Pfaffeneder, B. Hackner, W. Feiden, U. Schüller, T. Carell, H. A. Kretzschmar, *Int. J. Cancer* **2012**, *131*, 1577-1590. "Low Values of 5-Hydroxymethylcytosine (5hmC), the "Sixth Base", are Associated with Anaplasia in Human Brain Tumors".
6. C. Kaul, M. Müller, **M. Wagner**, S. Schneider, T. Carell, *Nat. Chem.* **2011**, *3*, 794-800. "Reversible Bond Formation Enables the Replication and Amplification of a Crosslinking Salen Complex as an Orthogonal Base Pair".
5. D. Globisch, D. Pearson, A. Hienzsch, T. Brückl, **M. Wagner**, I. Thoma, P. Thumbs, V. Reiter, A. C. Kneuttinger, M. Müller, S. A. Sieber, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 9739-9742. "Systems-Based Analysis of Modified tRNA Bases".

4. D. Pearson, A. Hienzsch, **M. Wagner**, D. Globisch, V. Reiter, D. Özden, T. Carell, *Chem. Comm.* **2011**, 47, 5196-5198. "LC-MS Based Quantification of 2 '-Ribosylated Nucleosides Ar(p) and Gr(p) in tRNA".
3. D. Globisch, M. Münzel, M. Müller, S. Michalakis, **M. Wagner**, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS ONE* **2010**, 5:e15367. "Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates".
2. M. Münzel, D. Globisch, T. Brückl, **M. Wagner**, V. Welzmler, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2010**, 49, 5375-5377. "Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain".
1. T. Brückl, D. Globisch, **M. Wagner**, M. Müller, T. Carell, *Angew. Chem. Int. Ed.* **2009**, 48, 7932-7934. "Parallel Isotope-Based Quantification of Modified tRNA Nucleosides".

Summary

Just like Einstein's famous equation $E = mc^2$ made its way into popular culture, the canonical bases G, C, A and T (or U) have been firmly established in common knowledge. However, even many scientists do not know that next to those canonical bases more than one hundred modified nucleosides are naturally occurring in nucleic acids. Most of these modifications are RNA nucleosides, with transfer-RNA (tRNA) showing the highest modification density among all RNA species. Here, the modified nucleosides are essential for efficiency and accuracy of translation.

While particular modifications have been studied in detail and their biological functions are well understood, little is known about the systemic behavior and interplay of a tRNA nucleoside modification ensemble as a whole. One reason for this has been the lack of methods allowing the exact and parallel quantification of modified RNA-nucleosides. Aware of such limitation, Thomas Carell and his co-workers Daniel Globisch and Tobias Brückl developed such a quantification method for RNA modifications. Using this mass spectrometry-based isotope dilution method, absolute amounts of modified RNA nucleosides can be determined in parallel with high precision.

By applying this method and with the help of Caterina Brandmayr, Tobias Brückl, and others, I examined the modification density of tRNA sets from different mouse and pig organs. For this reason, a representative selection of modified tRNA nucleosides within each organ was quantified. Using these data, we could show that the overall modification content varies substantially between the tissues, with each organ possessing a characteristic modification density in its tRNA set. Within a certain tRNA set, all of the selected modifications show the same relative abundance in comparison to the respective modifications in tRNA sets of other organs. Furthermore, the observed tRNA modification densities correlate positively with organ-specific demands for protein synthesis. Interestingly, our measurements also revealed slight differences between mouse and pig in their relative tRNA modification density across organs. Taken together, these results hint at a regulation of protein synthesis efficiency in tissues by modulation of the tRNA modification density.

Continuing the investigation of tRNA modifications, I analyzed the levels of these nucleosides in human cancer cell lines. I could show that each cancer cell line investigated had its characteristic set of modifications, which in principle could be used for cell line identification. Further statistical analysis proved that the variance of modification density within a certain tRNA ensemble is higher in cancer cell lines compared to primary cell lines.

While more than one hundred different RNA modifications are known to date, only six enzymatically generated DNA modifications have been identified in mammals until now. With the exception of 5-methylcytosine (mC), all of those bases or nucleosides were found after I started my PhD.

The modified cytosine base mC and the corresponding DNA-methyltransferases (DNMTs) are key players in epigenetic gene regulation. For instance, methylation of CpG-dinucleotides in the promotor region of a gene generally suppresses its transcription. Accordingly, promotor demethylation leads to the expression of formerly “switched-off” genes. Since such demethylation events are also known to occur in non-dividing cells like post-mitotic neurons, DNA demethylation in these cases has to proceed actively and independent from replication. Several mechanisms have been proposed for active DNA demethylation, but until 2009, none of them could ultimately be proven as correct. With respect to the efforts of understanding the mechanism of active DNA demethylation, identification of the TET enzymes as a new family of DNA-modifying enzymes was a game-changing event for the scientific community. TET enzymes are capable of oxidizing mC to 5-hydroxymethylcytosine (hmC), and they also generate the newly discovered DNA bases 5-formylcytosine (fC) and 5-carboxycytosine (caC) by further oxidation of hmC. To elucidate the biological role of hmC, fC and caC, we adapted our LC-MS–based isotope-dilution method previously developed for the investigation of modified RNA nucleosides for exact and parallel quantification of modified DNA nucleosides.

To generate a solid data set as a starting point for further investigations, Daniel Globisch and Martin Münzel determined the absolute levels of hmC in various mouse organs, including different brain regions. Susanne Koch and I were responsible for preparing the DNA needed for these studies. We could show that all organs under investigation feature the same level of mC, which is 4.5% with respect to all cytosine bases and cytosine modifications. In contrast, the hmC level varies strongly between the organs: while the highest levels of hmC are found in brain tissues (0.3–0.7%), other organs only show an hmC level of around 0.2%.

Continuing the analysis of mouse DNA, Jessica Steinbacher and I could detect the presence of the newly discovered fC-base in different organs of adult mice. Interestingly, we observed a positive correlation between the levels of fC and hmC, meaning that brain tissues also feature the highest levels of fC. However, comparing the absolute levels of fC and hmC in mouse organs, fC is about ten times less abundant than hmC.

We then investigated age-dependent changes of hmC and fC content in mammals and found a positive correlation between hmC-levels and age in brain tissues of human individuals, as previously observed in mouse. However, in human brain tissues the maximum levels of hmC are considerably higher than in mouse brain. Additionally, I could show that the levels of fC in mammalian organs are dramatically dropping during early postnatal developmental stages despite the simultaneous increase in hmC levels. Our results are in good agreement with widely accepted hypotheses about these cytosine modifications: while hmC, like mC, seems to be a relatively stable epigenetic information carrier in DNA, fC (and caC) seem to predominantly act as intermediates of an active DNA

demethylation process. This is highlighted by the fact that DNA glycosylase TDG is capable of excising both fC and caC from DNA. In the course of the base excision repair process, the resulting abasic sites are then replaced by unmodified cytosine bases.

Until this mechanism was widely accepted, deamination of hmC to 5-hydroxymethyluracil (hmU) was proposed as being a key step of active DNA demethylation. However, working together with Toni Pfaffeneder, Fabio Spada, Caterina Brandmayr, and others, I could show that in tissues of adult mice no deamination of hmC takes place. Our cell culture experiments showed that hmU can instead be generated by TET-mediated oxidation of thymine bases in DNA.

In the course of our investigations on hmC and working together with Theo Kraus, Daniel Globisch, and others, I quantified the levels of this base in human brain tumor samples. We and other groups could show that the level of hmC is remarkably lower in tumor tissues compared to healthy ones. Furthermore, we found a negative correlation between WHO classification-based tumor malignancy and hmC-level.

I also worked on elucidating the biological and epigenetic role of these newly discovered DNA cytosine modifications. I could show that UHRF2, which was identified as an hmC-specific reader protein in a cooperation project with Michiel Vermeulen in Utrecht, is capable of influencing the catalytic activity of TET enzymes. Co-expression of UHRF2 and TET1 in HEK cells not only increases the level of hmC, but also leads to a dramatic increase in the fC and caC levels. In another project, Arshan Perera, David Eisen, and I examined the biological role of TET enzymes in the retina. Our findings suggest that the transcriptional inhibitory protein REST / NRSF is able to recruit TET3 for a context-specific hydroxylation of methylated DNA-sections, resulting in the expression of formerly silenced genes.

Zusammenfassung

Ähnlich wie Einsteins berühmte Formel $E = mc^2$ in die Populärkultur eingegangen ist, besitzen die kanonischen Basen G, C, A und T bzw. U eine Bekanntheit, die weit über naturwissenschaftlich interessierte Kreise hinausreicht. Dass neben den kanonischen Basen über hundert modifizierte Nukleoside ein weiterer natürlicher Bestandteil der Nukleinsäuren sind, ist jedoch weitgehend unbekannt. Die meisten dieser Modifikationen findet man in der RNA, wobei die Transfer-RNA (tRNA) die höchste Modifikationsdichte aller RNA-Spezies besitzt. Hier haben die Modifikationen einen direkten Einfluss auf die Effizienz und Genauigkeit der Translation.

Während einzelne tRNA-Modifikationen und ihre biologische Funktion sehr genau untersucht und verstanden sind, wurde das systemische Zusammenspiel der modifizierten Nukleoside einer tRNA oder eines tRNA-Sets als Gesamtheit bisher wenig erforscht. Ein Grund dafür war das Fehlen einer ausgereiften Methode zur parallelen Quantifizierung von modifizierten RNA-Nukleosiden. Eine solche Methode war deshalb im Arbeitskreis Carell von Daniel Globisch und Tobias Brückl entwickelt worden. Als massenspektrometrie-basierte Isotopenverdünnungsmethode erlaubt diese Methode eine exakte und parallele Quantifizierung der absoluten Nukleosidmengen in einer Probe.

Mit dieser Methode und in Zusammenarbeit mit Caterina Brandmayr, Tobias Brückl und anderen habe ich den Modifikationsgrad der tRNA-Sets in Mäuse- und Schweineorganen untersucht. Dafür wurde für jedes Organ eine repräsentative Auswahl an tRNA-Modifikationen exakt quantifiziert. Es stellte sich heraus, dass die aus den Organen isolierten tRNA-Sets jeweils eine unterschiedliche und für das Organ charakteristische Modifikationsdichte aufweisen. Dies gilt für praktisch alle untersuchten Modifikationen eines bestimmten tRNA-Sets im Vergleich zu den entsprechenden Modifikationen der tRNA-Sets anderer Organe. Die Modifikationsdichte des tRNA-Sets korreliert dabei positiv mit dem Proteinsynthesbedarf des Organs. Ordnet man die Organe nach ihrer tRNA-Modifikationsdichte, so findet man interessanterweise eine teilweise unterschiedliche Reihenfolge der Organe bei Mäusen und Schweinen und daneben bei Mäusen eine etwas höhere allgemeine Modifikationsdichte. Diese Ergebnisse könnten darauf hindeuten, dass in Säugetieren über den Modifikationsgrad des tRNA-Sets eine Regulation der Proteinsyntheserate möglich ist.

In einem weiteren Projekt habe ich die tRNA-Modifikationen in humanen Krebszelllinien untersucht. Ich konnte zeigen, dass jede Krebszelllinie ein für diese Zelllinie charakteristisches Modifikationsset besitzt, welches prinzipiell für eine Identifikation der Zelllinie genutzt werden kann. Wie die von mir durchgeführte statistische Analyse ergab, weist das tRNA-Set einer Krebszelllinie daneben im Vergleich zu primären Zelllinien eine größere Varianz der einzelnen Modifikationen auf.

Während eine kaum überschaubare Vielzahl an RNA-Modifikationen existiert, wurden in Säugetieren bisher nur fünf enzymatisch generierte DNA-Modifikationen identifiziert. Mit Ausnahme von 5-Methylcytosin (mC) wurden diese Basen bzw. Nukleoside erst während meiner Promotionszeit entdeckt.

Die Base mC und die für ihre Bildung verantwortlichen DNA-Methyltransferasen spielen eine Schlüsselrolle bei der epigenetischen Genregulation. So verhindert eine Methylierung der CpG-Dinukleotide in der Promotorregion eines Gens dessen Transkription. Lange Zeit wurde nach einem plausiblen Mechanismus für aktive, d.h. enzymatische und damit von der DNA-Replikation unabhängige DNA-Demethylierungsereignisse gesucht, wie sie z.B. in postmitotischen Neuronen auftreten. Die Entdeckung der TET-Enzyme, einer neuen Familie DNA-modifizierender Enzyme, wirkte daher elektrisierend. Diese Enzyme können mC zu 5-Hydroxymethylcytosin (hmC) oxidieren und sind durch Weiteroxidation von hmC auch für die Bildung der etwas später entdeckten DNA-Basen 5-Formyl- (fC) und 5-Carboxycytosin (caC) verantwortlich. Um die Frage nach der biologisch-epigenetischen Bedeutung dieser neuentdeckten Basen zu beantworten, wurde im Arbeitskreis Carell die für RNA-Nukleoside entwickelte Quantifizierungsmethode für die HPLC-MS- und UHPLC-MS/MS-basierte Quantifizierung von DNA-Modifikationen adaptiert und angewendet.

Um eine solide Basis für weitere Untersuchungen zu schaffen, haben Daniel Globisch und Martin Münzel zunächst die absoluten hmC-Mengen in verschiedenen Hirngeweben und Organen der Maus bestimmt. Die dafür benötigte DNA wurde von mir und Susanne Koch isoliert. Es zeigte sich, dass das mC-Level in den von uns untersuchten Organen, bezogen auf alle Cytosinbasen und Cytosinmodifikationen, gleichbleibend etwa 4,5% beträgt. Dagegen variiert das hmC-Level stark zwischen den Organen und Geweben. Die mit 0,3-0,7% höchsten hmC-Level wurden in Hirngeweben gemessen, während in den anderen Organen nur etwa 0,2% aller Cytosinbasen hydroxymethyliert vorliegen.

Zusammen mit Jessica Steinbacher konnte ich zeigen, dass auch die Base fC in allen Organen der adulten Maus vorkommt. Dabei korreliert das fC- mit dem hmC-Level, so dass auch die höchsten von uns detektierten fC-Mengen in den Hirngeweben gefunden wurden. Allerdings betragen die in der Maus gefundenen absoluten fC-Level nur jeweils etwa 10% des hmC-Levels.

Auch die Änderungen der hmC- und fC-Level in Abhängigkeit vom Lebensalter wurden von mir und Jessica Steinbacher untersucht. Dabei haben wir den bereits von der Maus bekannten Anstieg des hmC-Levels mit zunehmendem Alter auch in menschlichen Hirngeweben beobachtet. Hier allerdings finden sich deutlich höhere Spitzenwerte. In der gleichen Studie konnten wir zeigen, dass die fC-Level in den Organen direkt nach der Geburt und in den frühen Entwicklungsstadien bei gleichzeitig ansteigender DNA-Hydroxymethylierung stark abnehmen. Unsere Ergebnisse passen sehr gut zu der

inzwischen weitgehend akzeptierten und experimentell belegten Hypothese, dass hmC-Basen eine längere Verweildauer in der DNA besitzen und eine eigene epigenetisch-regulatorische Bedeutung für die Genexpression haben. Dagegen werden fC und caC im Wesentlichen als direkte Intermediate der aktiven DNA-Demethylierung angesehen. Sie können im Rahmen der Basenexzisionreparatur von der Glykosylase TDG aus der DNA herausgeschnitten und durch unmodifizierte Cytosine ersetzt werden. Bevor dieser Mechanismus allgemein akzeptiert wurde, ist eine Zeitlang vermutet worden, dass die Desaminierung von hmC zu 5-Hydroxymethyluracil (hmU) ein Schlüsselschritt der aktiven DNA-Demethylierung sein könnte. Zusammen mit Toni Pfaffeneder, Fabio Spada, Caterina Brandmayr und anderen habe ich zeigen können, dass in den Organen und Geweben der Maus praktisch keine hmC-Desaminierung stattfindet. Stattdessen konnten wir nachweisen, dass hmU durch eine TET-vermittelte Oxidation von Thymin gebildet werden kann.

In einem weiteren Projekt habe ich zusammen mit Theo Kraus, Daniel Globisch und anderen die hmC-Level in Tumorgeweben quantifiziert. Wir konnten –zusammen mit anderen Arbeitsgruppen– zeigen, dass Tumore im Vergleich zu gesundem Gewebe ein ungewöhnlich niedriges hmC-Level besitzen. Darüber hinaus konnten wir bei den von uns untersuchten Hirntumoren eine negative Korrelation zwischen hmC-Gehalt und WHO-Tumorgrad nachweisen.

Weiterhin habe ich auch an der Aufklärung der epigenetisch-biologischen Bedeutung der neuentdeckten Cytosinmodifikationen mitgearbeitet. So konnte ich nachweisen, dass UHRF2, welches in unserem Kooperationsprojekt mit Michiel Vermeulen in Utrecht als spezifischer hmC-Reader identifiziert wurde, die katalytischen Eigenschaften der TET-Enzyme beeinflussen kann. Eine von mir durchgeführte Coexpression von UHRF2 und TET1 resultierte nicht nur in einem Anstieg des detektierbaren hmC-Levels, sondern führte darüber hinaus auch zu einer starken Zunahme des fC- und caC-Levels. In einem weiteren Kooperationsprojekt wurde von Arshan Perera, David Eisen und mir die biologische Bedeutung der TET-Enzyme in der Retina untersucht. Die Ergebnisse dieses Projektes legen nahe, dass TET3 durch das transkriptionsinhibierende Protein REST / NRSF für eine kontextspezifische Hydroxylierung methylierter DNA-Abschnitte rekrutiert wird. Dies führt zu einer Expression zuvor stummgeschalteter Gene.

Inhaltsverzeichnis

1	Modifizierte Nukleoside als epigenetische Informationsträger	11
1.1	Modifizierte Nukleoside in RNA und DNA	11
1.2	Epigenetik	13
2	RNA-Epigenetik ?	16
2.1	Regulation der Translation durch die mRNA-Modifikation m ⁶ A	16
2.2	Existiert eine dynamische Regulation des gesamten tRNA-Modifikationssets?	19
3	Modifizierte DNA-Nukleoside: Vorkommen, biologische Bedeutung und Einfluss auf die Genexpression	20
3.1	Eine neue Familie DNA-modifizierender Enzyme	20
3.1.1	Die TET-Enzyme: Substrate und Produkte, katalysierte Reaktionen und Struktur	20
3.1.2	Mechanismus der TET-katalysierten Oxidation von Methylcytosin	22
3.1.3	Nichtenzymatische biologische Wirkungen der TET-Enzyme	23
3.1.4	Expressionslevel der TET-Enzyme und Regulation ihrer Aktivität	24
3.1.4.1	Expressionslevel der TET-Enzyme in Organen und Zelltypen	24
3.1.4.2	Regulation und Kontrolle der TET-Aktivität	25
3.1.5	TET-Mutanten, TET-Fehlregulation und damit einhergehende Erkrankungen	25
3.2	Globale Quantifizierung und genomische Sequenzierung der Basen hmC, fC, caC und hmU	27
3.2.1	Quantifizierung der globalen Cytosinmodifikationslevel in Organen und Geweben	27
3.2.2	Affinitätsanreicherungsverfahren	28
3.2.2.1	Methoden zur Affinitätsanreicherung von hmC	29
3.2.2.2	Methoden zur Affinitätsanreicherung von fC und hmU	30
3.2.3	Sequenzierungsmethoden mit basengenaue Auflösung	31
3.2.3.1	Basengenaue Sequenzierung von hmC	32
3.2.3.2	Basengenaue Sequenzierung von fC und caC	33
3.2.4	Sequenzierungsmethoden der dritten Generation	34
3.2.4.1	SMRT-Sequencing	34

3.2.4.2	Nanopore-Sequencing	35
3.3	Vorkommen und Verteilung von hmC, fC, caC und hmU im Genom	37
3.3.1	5-Hydroxymethylcytosin (hmC).....	37
3.3.2	5-Formylcytosin (fC) und 5-Carboxycytosin (caC)	39
3.3.3	5-Hydroxymethyluracil (hmU).....	42
3.4	DNA-Demethylierung in Säugetieren	43
3.4.1	Globale und lokale DNA-Demethylierungen in Säugetieren	44
3.4.1.1	Globale DNA-Demethylierungen während der Embryogenese	44
3.4.1.2	Lokusspezifische, aktive DNA-Demethylierung in somatischen Zellen	46
3.4.2	Mögliche Mechanismen der aktiven DNA-Demethylierung.....	48
3.4.2.1	Aktive DNA-Demethylierung über die Glykosylase TDG im Rahmen der Basenexzisionsreparatur.....	50
3.4.2.2	Demethylierung via Desaminierung.....	51
3.4.2.3	Aktive DNA-Demethylierung durch enzymatische Eliminierung der oxidierten Methylgruppe	54
3.5	hmC-, fC- und caC-bindende Proteine als Übersetzer der epigenetischen Information.....	57
3.5.1	Einfluss von Cytosinmodifikationen in der DNA auf die Bindungsaffinität DNA-bindender Proteine zu ihren Zielsequenzen	57
3.5.2	Massenspektrometriebasierte Identifikation von hmC-, fC-, und caC-bindenden Proteinen	58
4	Ergebnisse und Publikationen.....	61
4.1	“Parallel Isotope-Based Quantification of Modified tRNA Nucleosides”	62
4.2	“Isotope-Based Analysis of Modified tRNA Nucleosides Correlates Modification Density with translational Efficiency”	66
4.3	Statistische Analyse der tRNA-Modifikationssets von Krebszelllinien als Basis für eine Klassifizierung	71
4.3.1	Vorarbeiten	72
4.3.2	Zur Klassifizierung von Krebszelllinien	73
4.3.3	Quantifizierung von tRNA-Modifikationen humaner Krebszelllinien.....	74
4.3.3.1	Auswahl der Zelllinien	75
4.3.3.2	Auswahl der tRNA-Modifikationen	76

4.3.3.3	Extraktion der tRNA und Quantifizierung der Modifikationen	77
4.3.3.4	Ergebnisse der Quantifizierung und Diskussion	78
4.3.4	Statistische Analyse der tRNA-Modifikationssets	82
4.3.4.1	Die Pearson-Korrelation und der Pearson-Korrelationskoeffizient als Ähnlichkeitsmaß zwischen Zahlen- bzw. Datenreihen	82
4.3.4.2	Normierung der Modifikationslevel	83
4.3.4.3	Durchführung der Clusteranalyse	83
4.3.4.4	Ergebnisse der Clusteranalyse	84
4.3.4.5	Validierung der Clusteranalyse	88
4.3.5	Zusammenfassung und Ausblick	95
4.3.6	Anhang; experimenteller Teil	97
4.3.6.1	Zellkultur, Extraktion der tRNA, enzymatischer Verdau und LC-MS-Messungen	97
4.3.6.2	Normierung der Nukleosidlevel für die Clusteranalysen	97
4.3.6.3	Für die Clusteranalyse der Modifikationssets der Krebszelllinien in <i>Cluster</i> gewählte Programmeinstellungen	98
4.4	“Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain”	100
4.5	“Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates”	104
4.6	Postnatale, altersabhängige Änderungen der DNA-Modifikationslevel in Säugetieren	114
4.6.1	Altersabhängige Variation der hmC- und mC-Level im menschlichen Hirn	116
4.6.1.1	Dynamik der Hydroxymethylcytosinlevel im menschlichen Hirn	118
4.6.1.2	Dynamik der Methylcytosinlevel im menschlichen Hirn	119
4.6.2	Altersabhängige Änderungen der Nukleosidlevel in Mäusegeweben	120
4.6.2.1	Altersabhängige Änderungen der Nukleosidlevel in Cerebrum und Niere	121
4.6.2.2	Detaillierte Untersuchung der altersabhängigen Änderungen der Nukleosidlevel in der Retina	125
4.6.2.3	Analyse der in der Retina gemessenen Nukleosidlevel anhand der Expressionsdaten wichtiger DNA-modifizierender Enzyme	127
4.6.3	Zusammenfassung und Diskussion	129
4.6.4	Anhang	132
4.7	“Age-Dependent Levels of 5-Methyl-, 5-Hydroxymethyl-, and 5-Formylcytosine in Human and Mouse Brain Tissues”	133

4.8	“Low Values of 5-Hydroxymethylcytosine (5hmC), the “Sixth Base”, are Associated with Anaplasia in Human Brain Tumors”	138
4.9	“Tet Oxidizes Thymine to 5-Hydroxymethyluracil in Mouse Embryonic Stem Cell DNA”	153
4.10	“Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives”	162
4.11	“TET3 is Recruitet by REST for Context-Specific Hydroxymethylation and Induction of Gene Expression”	177
5	Abkürzungsverzeichnis	190
6	Literaturverzeichnis.....	192

1 Modifizierte Nukleoside als epigenetische Informationsträger

1.1 Modifizierte Nukleoside in RNA und DNA

Zum Grundwissen unserer Zeit gehört die Kenntnis der Grundbestandteile unseres Erbguts. Tatsächlich sind A, G, C und T bzw. U mittlerweile feste Bestandteile der Alltagskultur, so dass Teile des frisch entzifferten genetischen Codes des Menschen sogar im Feuilleton einer bekannten Tageszeitung abgedruckt wurden.^[1] Kaum bekannt ist dagegen, dass diese kanonischen Basen nur einen kleinen Teil der gesamten chemischen Vielfalt von RNA und DNA ausmachen.^[2] Besonders in der RNA findet man eine Vielzahl von modifizierten Nukleosiden, welche oftmals entscheidend für ihre biologische Funktion sind. Mittlerweile kennt man mehr als hundert nicht-kanonische RNA-Nukleoside, deren strukturelle Diversität von einfachen Methylierungen^[3] bis zu hypermodifizierten Nukleosiden wie z.B. Wybutosin^[4] reicht (siehe Abbildung 1). Diese modifizierten Nukleoside sind in praktisch allen RNA-Spezies zu finden, wobei snRNA, mRNA, rRNA und tRNA teilweise spezifische Sets von Modifikationen besitzen, welche sich zudem zwischen den Organismen unterscheiden.^[5-7]

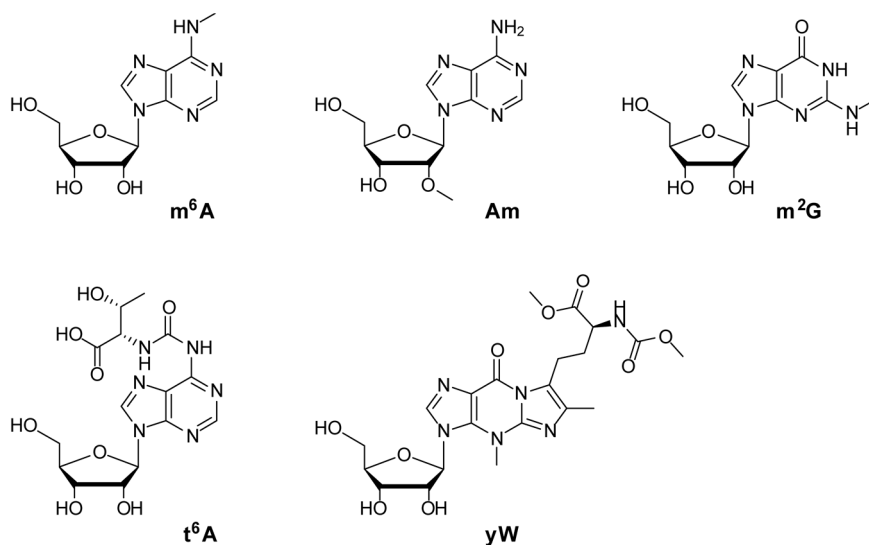


Abbildung 1. Beispiele für natürlich vorkommende nichtkanonische RNA-Nukleoside: m⁶A (N⁶-Methyladenosin); Am (2'-O-Methyladenosin); m²G (N²-Methylguanosin); t⁶A (N⁶-Threonylcarbamoyladenosin); yW (Wybutosin).

Die höchste Modifikationsdichte beobachtet man bei der tRNA.^[8] Als Adaptermolekül ist diese tRNA-Spezies direkt an der Übersetzung des genetischen Codes in Proteine, der sogenannten Translation, beteiligt. So wird im Ribosom für jedes Basentriplett (Codon) auf der mRNA entsprechend dem genetischen Code die passende Aminosäure von einem tRNA-Molekül auf die wachsende Peptidkette

übertragen.^[9] Dabei gibt es für jede Aminosäure mindestens eine, meistens jedoch mehrere spezifische tRNAs, welche über ihre Anticodonregion mit der mRNA wechselwirken.^[10] Entscheidend für eine fehlerfreie Translation ist die korrekte Paarung zwischen mRNA-Codon und tRNA-Anticodon. Die Spezifität dieser Wechselwirkung wird unter anderem durch modifizierte Nukleoside in der Anticodonregion der tRNA sichergestellt.^[11] Andere tRNA-Modifikationen sorgen für eine korrekte Faltung der tRNA,^[12] stabilisieren die Tertiärstruktur^[13] oder bilden die strukturelle Grundlage für die substratspezifische Bindung zwischen der tRNA und ihrer Aminoacyl-tRNA-Synthetase.^[14]

Weitaus weniger modifizierte Nukleoside findet man dagegen in der DNA. Lange Zeit war 5-Methylcytosin (mC) die einzige bekannte DNA-Modifikation in Säugetieren.^[15] mC wird von den DNA-Methyltransferasen (DNMTs) gebildet, welche unmodifizierte Cytosinbasen in Position 5 methylieren können.^[16-17] In Säugetieren ist Methylcytosin die häufigste DNA-Modifikation, hier sind etwa 4,5% aller Cytosine methyliert.^[18] Methylierungen sind besonders oft im palindromischen CpG-Dinukleotid- Sequenzkontext zu finden.

Die Base Methylcytosin ist für die langfristige Fixierung des Transkriptionszustands einzelner Gene auf DNA-Ebene von größter Bedeutung. Der Methylierungsgrad der CpG-Dinukleotide in der Promotorregion eines Gens beeinflusst dessen Transkription in charakteristischer Weise: So sind Promotoren von aktiv transkribierten Genen häufig hypomethyliert, d.h. sie enthalten nur wenige mC-Basen.^[19-21] Dagegen führt eine Methylierung der CpG-Dinukleotide des Promotors zu einer dauerhaften Stummschaltung des assoziierten Gens, welche sogar auf die nächste Generation vererbt werden kann.^[19-21] Das DNA-Methylierungsmuster (d.h. das Methylom) einer Zelle legt daher genau fest, welche Gene in dieser Zelle exprimiert werden und welche nicht. Da jeder Zelltyp ein eigenes, charakteristisches Methylom aufweist, werden nur zelltypspezifische Gene in den entsprechenden Zellen exprimiert, obwohl alle Zellen des Organismus dasselbe Genom besitzen.

In den letzten Jahren wurden mit 5-Hydroxymethylcytosin (hmC),^[22-24] 5-Formylcytosin (fC)^[25-26] und 5-Carboxycytosin (caC)^[26-27] weitere, bis dato unbekannte Cytosinderivate in der DNA von Säugetieren entdeckt. Sie werden von den TET-Enzymen, einer Familie von α -Ketoglutarat abhängigen Dioxygenasen, durch Oxidation von Methylcytosin gebildet.^[28-30] Wie Carell und Mitarbeiter zeigen konnten, wird auch die Base 5-Hydroxymethyluracil (hmU) in vergleichbarer Weise durch TET-vermittelte Oxidation von Thymin (T) gebildet.^[31] hmU war zuvor nur als oxidativer DNA-Schaden bekannt (siehe Abbildung 2).^[32]

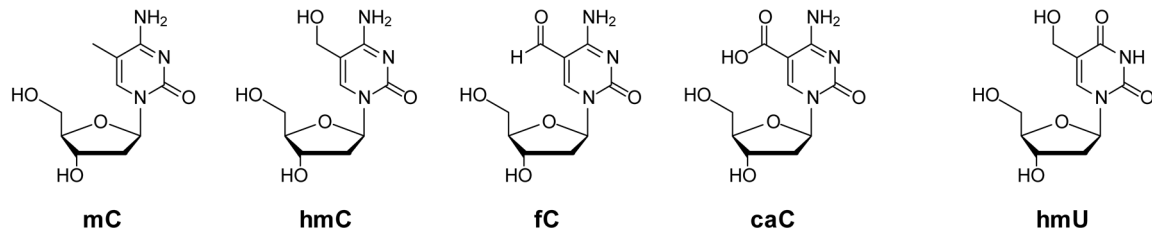


Abbildung 2. Enzymatisch generierte nichtkanonische DNA-Nukleoside bzw. -Basen. Neben den Cytosinmodifikationen mC (5-Methylcytosin), hmC (5-Hydroxymethylcytosin), fC (5-Formylcytosin) und caC (5-Carboxycytosin) findet man auch die Base hmU (5-Hydroxymethyluracil).

Mittlerweile weiß man, dass die neuentdeckten Basen hmC, fC und caC als Intermediate an der aktiven, d.h. enzymatischen und damit replikationsunabhängigen Demethylierung der DNA beteiligt sind.^[18, 33-35] Dabei werden methylierte wieder in unmodifizierte Cytosine überführt. Der exakte Mechanismus dieser aktiven Demethylierung wird derzeit intensiv untersucht.^[36] Aktive DNA-Demethylierungsprozesse ermöglichen der Zelle eine Adaption des Transkriptionsniveaus einzelner Gene und damit eine situativ angepasste Genexpression. Dies ist besonders während der Embryogenese^[37-38] und für die Aufrechterhaltung der neuronalen Plastizität im adulten Tier^[39] von außerordentlicher Bedeutung. Es mehren sich zudem die Hinweise darauf, dass besonders hmC nicht nur an der aktiven DNA-Demethylierung beteiligt ist, sondern darüber hinaus, ähnlich wie mC, eine Funktion als epigenetischer Informationsträger besitzt.^[40-41]

1.2 Epigenetik

Die vollständige Entzifferung des menschlichen Genoms im Jahre 2001 war ein Triumph der modernen Molekularbiologie und ein Meilenstein auf dem Weg zum besseren Verständnis der menschlichen Natur.^[42-43] Allerdings kann aus dem genetischen Code allein, d.h. aus der Abfolge der kanonischen Watson- und Crick-Basen in der DNA nicht auf das Proteom einer Zelle, d.h. auf die zu einem bestimmten Zeitpunkt in dieser Zelle exprimierten Proteine, geschlossen werden. Dies ist nicht überraschend, wenn man bedenkt, dass jede Zelle eines Organismus das gleiche Genom enthält. Dennoch unterscheidet sich z.B. eine Hautzelle erheblich von einer Nervenzelle. Diese Unterschiede finden sich auch auf molekularer Ebene wieder und sind das Resultat einer zelltypspezifischen Genexpression. Wie kommt es aber, dass nur ein Teil aller vorhandenen Gene der Zelle exprimiert wird, oder anders gefragt, warum existieren zu einem bestimmten Genotyp unterschiedliche Phänotypen? Es müssen offenbar Mechanismen in einer Zelle existieren, welche über die Basenabfolge der DNA hinaus eine Regulation der Genexpression erlauben. Für diese Regulation existieren verschiedene Mechanismen. Kurzfristige Änderungen der Transkription und damit auch der

Genexpression werden meistens über Transkriptionsfaktoren gesteuert. Dabei handelt es sich um DNA-bindende Proteine, welche z.B. die RNA-Polymerase II oder histonmodifizierende Enzyme rekrutieren können und auf diese Weise die Transkription eines Gens beeinflussen. Daneben gibt es Mechanismen, welche eine mittel- oder langfristige Fixierung der Aktivität eines Gens bzw. eines zellulären Genexpressionsprofils erlauben. Mit der Untersuchung und Aufklärung dieser von der primären Basensequenz der DNA unabhängigen Mechanismen beschäftigt sich die Epigenetik.^[44-47]

Tatsächlich wird der Begriff *Epigenetik* manchmal fälschlicherweise für jede Art der Genregulation verwendet. Bei einer konservativen Auslegung des Begriffs muss sogar eine Vererbbarkeit des Genexpressionsprofils an die Folgegenerationen möglich sein. Erst dann darf man einen Prozess, welcher die Genaktivität steuert, als epigenetisch bezeichnen. In dieser Arbeit wird der Begriff eher progressiv ausgelegt und verwendet, um alle Prozesse zu beschreiben, welche durch posttranslationale Modifikation von Biomolekülen die Aktivität eines Gens bzw. das zelluläre Genexpressionsprofil mittelfristig oder langfristig in einem bestimmten Zustand fixieren.

Enzyme, welche an einer epigenetischen Regulation der Genexpression direkt beteiligt sind, treten oft in der Rolle von *Writer*, *Reader* oder *Eraser* auf. Während Writer epigenetisch wirkende chemische Modifikationen in eine Nukleinsäure bzw. in ein Protein einführen, erkennen Reader diese Modifikationen und übersetzen sie in eine biologische Wirkung. Eraser können epigenetische Modifikationen wieder aus der Nukleinsäure bzw. aus dem Protein entfernen. Das Zusammenspiel von Writern, Readern und Erasern ermöglicht einen vollständigen ON-OFF-ON bzw. OFF-ON-OFF-Zyklus, bei dem ein Gen je nach biologischem Bedarf exprimiert oder nicht exprimiert wird (siehe Abbildung 3).^[48]

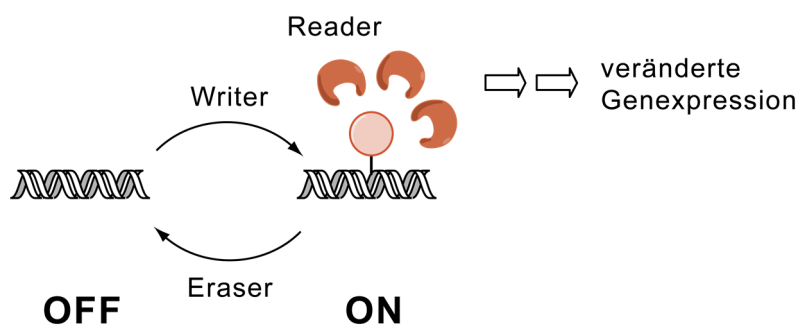


Abbildung 3. Schematische Darstellung eines dynamischen epigenetischen Systems am Beispiel einer beliebigen DNA-Modifikation. Die Zuordnung von *OFF* und *ON* zu den beiden Schaltzuständen ist willkürlich gewählt.

In höheren Organismen existieren mehrere Wege, um die Aktivität eines Gens bzw. das zelluläre Genexpressionsprofil mittel- oder langfristig zu fixieren. So kann die Transkription durch chemische Modifizierung der Cytosinbasen innerhalb bestimmter DNA-Abschnitte direkt und dauerhaft reguliert werden.^[49] Daneben kann die Genaktivität über die lokale Packungsdichte der DNA gesteuert werden, da für eine aktive Transkription der freie Zugang der beteiligten Enzyme zur DNA notwendig ist.^[50] Im Zellkern ist die kleinste DNA-Packungseinheit das Nukleosom. Dabei handelt es sich um ein Oktamer aus acht Histonen, welches in 1,65 Windungen von DNA umgeben ist.^[51] Die Histone besitzen am N-terminalen Ende einen flexiblen Fortsatz, dessen Aminosäuren von Enzymen modifiziert werden können. Dabei können neben Methylierungen, Acetylierungen, und Phosphorylierungen auch komplexere Modifikationen wie Ubiquitynlierungen und Sumoylierungen auftreten.^[52-53] Dieses komplexe Muster von Histonmodifikationen dient als Code, mit dessen Hilfe die Packungsdichte der DNA und damit die lokale Transkriptionsaktivität im Zellkern gesteuert werden kann.^[54-55] Die chemische Modifizierung der Cytosinbasen einerseits und der Histone andererseits sind keine unabhängigen epigenetischen Prozesse, sondern sind auf vielfältige Weise miteinander verwoben.^[54, 56-58]

Sowohl im Falle der Histon- als auch bei der Cytosinmodifizierung führt eine durch Enzyme vermittelte chemische Modifikation von Aminosäuren bzw. DNA-Basen zu einer mittel- oder langfristigen Änderung der Genexpression. Erst seit kurzem weiß man, dass in Säugetieren ein mRNA-basiertes System zur Steuerung der Genexpression existiert, welches auch auf einer reversiblen chemischen Basenmodifikation beruht.^[59] Hier ist eine enzymkatalysierte Methylierung der Aminogruppe von Adenin (A), bei der das RNA-Nukleosid *N*⁶-Methyl-Adenosin (*m*⁶A) gebildet wird, die Basis für eine Regulation der mRNA-Translation und damit auch der Genexpression.^[48, 60-61] Wegen der geringen Halbwertszeit der mRNA in der Zelle handelt es sich hierbei um einen Mechanismus zur kurzfristigen Expressionsregulation und damit nicht um ein klassisches epigenetisches System.

In den folgenden Kapiteln wird zunächst eine kurze Einführung in die Regulation der Genexpression auf RNA-Ebene gegeben. Danach wird die epigenetische Steuerung der Genexpression durch DNA-Cytosinmodifizierung genauer beschrieben. Da die Histonmodifizierung und Histonmodifikationen nicht Teil dieser Arbeit waren, wird dieser epigenetische Mechanismus nur kurz im Zusammenhang mit der Cytosinmodifizierung angesprochen.

2 RNA-Epigenetik ?

2.1 Regulation der Translation durch die mRNA-Modifikation m⁶A

Obwohl die Regulation der Genexpression durch DNA-Cytosinmethylierung seit Jahrzehnten bekannt ist (siehe Abschnitt 3), war es eine große Überraschung, dass ein ähnlicher Mechanismus auch auf mRNA-Ebene existiert.^[48, 60, 62] Auch hier führt eine selektive, enzymkatalysierte Basenmodifizierung zu einer veränderten Genexpression.^[61] Im Falle der mRNA ist nicht Methylcytosin, sondern die Base N⁶-Methyladenin bzw. das Nukleosid N⁶-Methyladenosin (m⁶A) der epigenetische Informationsträger (siehe Abbildung 4).^[48, 60, 63-66] m⁶A ist die häufigste mRNA-Modifikation; pro RNA-Strang sind durchschnittlich drei bis vier m⁶A-Modifikationen vorhanden.^[67-69] Während modifizierte DNA-Basen die Genexpression über eine Modulation der Transkription kontrollieren, steuert m⁶A die Genexpression durch Beeinflussung der Translation.^[60] Es handelt sich hier um ein dynamisches System, d.h. neben Methyltransferasen als m⁶A-Writer gibt es auch Eraser-Enzyme, welche diese Modifikation wieder in A umwandeln können (siehe Abbildung 4).^[48] Dies ermöglicht eine flexible Reaktion des m⁶A-basierten Regulationssystems auf äußere Reize nach dem OFF-ON-OFF bzw. ON-OFF-ON-Prinzip.

Natürlich handelt es sich hier nicht um ein „klassisches“ epigenetisches System. Aufgrund der geringen Halbwertszeit der mRNA in der Zelle ist eine mittel- bzw. langfristige m⁶A-basierte Genregulation gar nicht möglich. Möglicherweise spielen jedoch ganz ähnliche Regulationsprozesse bei der Steuerung der Translation durch tRNA-Modifizierung eine Rolle (siehe Abschnitt 2.2). Außerdem weist die m⁶A-basierte Translationsregulation, abgesehen vom zeitlichen Aspekt, alle typischen Merkmale eines epigenetischen Systems auf, so wie man es klassischerweise z.B. von der DNA-Methylierung kennt. Es bietet daher einen idealen Einstieg in dieses Themengebiet.

Den ersten Hinweis auf die besondere biologische Bedeutung von m⁶A lieferte die Beobachtung, dass ein Knockdown oder Knockout der katalytischen Untereinheiten des für die Bildung von m⁶A verantwortlichen RNA-Methyltransferaskomplexes zum Zelltod führt.^[70] Dieser Komplex besteht aus den Proteinen METTL3,^[70] METTL14^[71] (METTL = Methyltransferase-like) und WTAP.^[71-72] Bei den METTL-Enzymen handelt es sich um die katalytischen Untereinheiten des Komplexes.^[71] Beide katalysieren für sich genommen die Bildung von m⁶A durch Methylierung der N⁶-Aminogruppe von Adenin und können als *Writer* vermutlich unterschiedliche Sets von potentiellen m⁶A-Positionen methylieren. WTAP besitzt dagegen keine eigene Methyltransferaseaktivität; ein Knockdown von WTAP führt jedoch zu einem signifikanten Rückgang von m⁶A in der mRNA.^[71] Es wird deshalb

vermutet, dass WTAP entweder die subnukleare Lokalisation des Komplexes steuert oder die Selektivität der Methyltransferaseuntereinheiten beeinflusst.^[72]

Für die Umwandlung von m^6A zurück zu A sind die m^6A -spezifischen Demethylasen FTO^[59] und ALKBH5^[73] als *Eraser* verantwortlich.^[74] Bei beiden Enzymen handelt es sich um Fe^{II} - und α -Ketoglutarat-abhängige Dioxygenasen.^[75-77] FTO und ALKBH5 gehören daher zur selben Enzymfamilie wie die TET-Enzyme, welche die Oxidation der DNA-Base mC zu hmC, fC und caC katalysieren (siehe folgende Kapitel).^[29, 78] Im Falle von m^6A wird die N^6 -Methylgruppe in zwei Schritten zunächst zu einem Hydroxyl- und dann zu einem Formylrest oxidiert (für Einzelheiten des Mechanismus siehe Abbildung 6).^[73, 79] Die dann wieder unmodifizierte Adeninbase entsteht durch anschließende Hydrolyse der Formamidgruppe. Interessanterweise scheint es einen wichtigen Unterschied zwischen FTO und ALKBH5 zu geben. So finden die beiden Oxidationsschritte bei der Demethylierung von m^6A durch ALKBH5 offenbar direkt hintereinander statt, d.h. ohne dass die Zwischenprodukte N^6 -Hydroxymethyladenin (hm^6A) und N^6 -Formyladenin (f^6A) als freie mRNA-Basen auftreten.^[73] Dagegen scheint die Oxidation von m^6A durch FTO sukzessive zu erfolgen; die Basen hm^6A und f^6A sollen dabei als eigenständige Modifikationen mit einer Lebensdauer von mehreren Stunden in der mRNA auftreten, bevor eine Weiteroxidation bzw. Hydrolyse erfolgt (eine schematische Darstellung des m^6A -basierten Systems findet sich in Abbildung 4).^[79]

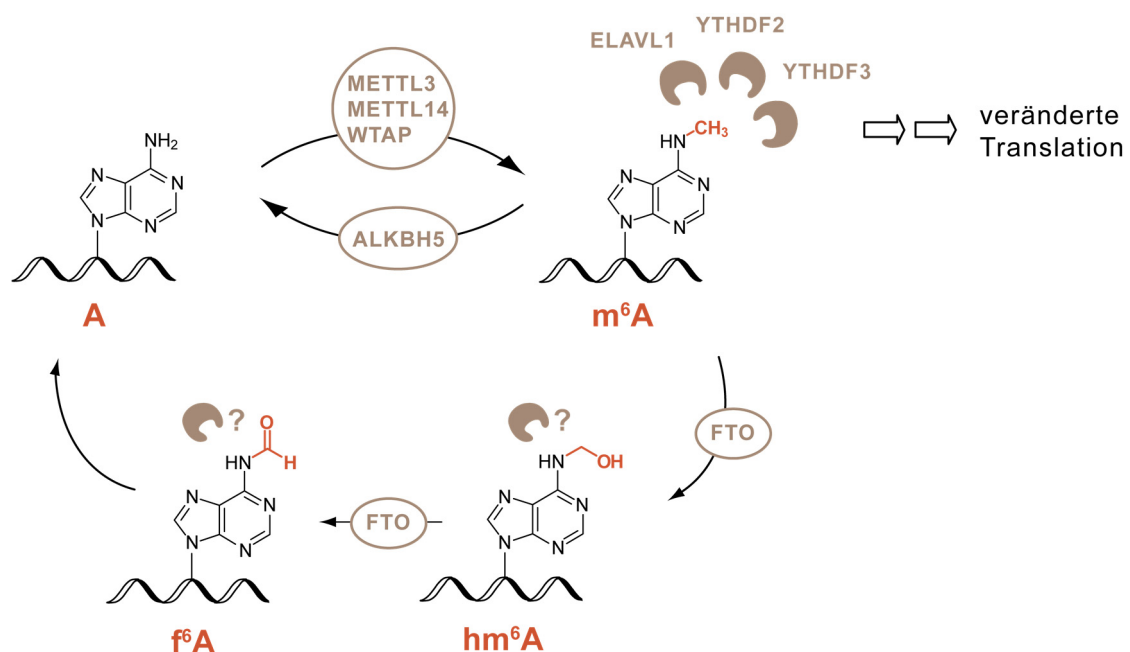


Abbildung 4. Schematische Darstellung des m^6A / mRNA-basierten Systems zur Kontrolle der Translation und damit der Genexpression. Die beteiligten Enzyme sind in hellbraun dargestellt. *Writer*: Komplex aus METTL3, METTL14 und WTAP; *Reader*: ELAVL1, YTHDF2 und YTHDF3; *Eraser*: ALKBH5 und FTO. f^6A (und evtl. auch hm^6A , hier nicht dargestellt,) werden ohne Enzymbeteiligung wieder zu A hydrolysiert.

Wie kann nun das auf m⁶A beruhende regulatorische System die Genexpression beeinflussen, d.h. wie wird die von m⁶A kodierte Information in eine biologische Wirkung übersetzt? Offenbar müssen Proteine existieren, welche als spezifische *Reader* methylierte Adeninbasen in der mRNA erkennen und binden können.^[80] Von Rechavi und Mitarbeitern konnten mehrere dieser m⁶A-spezifischen Interaktoren identifiziert werden.^[81] Dabei handelt es sich um die Proteine ELAVL1 bzw. HuR1, YTHDF2 und YTHDF3. Alle drei Proteine beeinflussen die mRNA-Stabilität in spezifischer Weise. So ist bekannt, dass die Proteine der ELAVL-Familie bevorzugt an AU-reiche Elemente innerhalb der mRNA binden,^[82] wobei mRNAs durch die Interaktion mit ELAVL1 stabilisiert werden.^[83] Weiterhin konnte gezeigt werden, dass mRNA-ELAVL1-Interaktionen in Stammzellen direkt vom Vorhandensein von m⁶A in der mRNA abhängen.^[84] Die m⁶A-Reader YTHDF2 und YTHDF3 gehören zu einer Superfamilie von Proteinen, welche alle die in Eukaryoten konservierte, RNA-bindende YTH-Domäne besitzen.^[85] YTHDF2 besitzt *in vitro* eine ca. 15-fach höhere Bindungsaffinität zu m⁶A-modifizierten RNA-Oligonukleotiden gegenüber unmodifizierten Oligonukleotiden.^[86] Im Zellplasma konkurriert YTHDF2 direkt mit den Ribosomen um m⁶A-modifizierte mRNAs.^[86] Solange genügend freie Ribosomen für eine Bindung zur Verfügung stehen, kann eine Translation der modifizierten mRNAs stattfinden. Je mehr Bindungsstellen an den Ribosomen belegt sind, umso größer ist der Anteil an m⁶A-modifizierten mRNAs, die stattdessen von YTHDF2 gebunden werden. Eine solche Bindung hat den Abbau dieser mRNAs zur Folge, u.a. indem sie zu den *P-Bodies*, den Orten der zellulären mRNA-Degradation, transportiert werden.^[86]

Besonders interessant ist die Entdeckung des m⁶A-basierten regulatorischen Systems auch deshalb, weil es an der Regulation des zellulären Energiestoffwechsels beteiligt zu sein scheint. So führt ein FTO-knockout bei Mäusen zu einem deutlich verminderten Körpergewicht.^[87] Beim Menschen wurde eine stark positive Korrelation gefunden zwischen dem Auftreten bestimmter FTO-Varianten und der Wahrscheinlichkeit, einen zu hohen *Body-Mass-Index* zu besitzen^[88-89] oder an Typ 2-Diabetes zu erkranken.^[90-91] Ein detailliertes Verständnis des m⁶A-Systems könnte daher möglicherweise der Ausgangspunkt für die Entwicklung von neuen Behandlungsmöglichkeiten der Volksleiden Übergewicht und Diabetes sein.

Bis jetzt ist m⁶A die einzige bekannte mRNA-Modifikation, deren Vorkommen dynamisch reguliert ist. Allerdings gibt es Hinweise darauf, dass auch das Auftreten von Pseudouridin (Ψ) in der mRNA einer Regulation unterworfen ist,^[92-94] obwohl in diesem Fall bisher weder *Reader* noch *Eraser* identifiziert werden konnten.^[95-96] In den kommenden Jahren wird die mRNA-basierte epigenetische Kontrolle der Genexpression im Fokus des Interesses liegen, und man darf gespannt sein, welche weiteren mRNA-Modifikationen an diesen Prozessen beteiligt sind.

2.2 Existiert eine dynamische Regulation des gesamten tRNA-Modifikationssets?

Neben der Kontrolle der Genexpression durch einzelne, dynamisch regulierte RNA-Modifikationen^[48] ist auch eine Steuerung dieser Prozesse auf höherer Ebene denkbar. Da besonders in der tRNA viele Modifikationen einen direkten Einfluss auf die Effizienz der Translation und damit der Genexpression haben,^[5] ist z.B. eine Kontrolle der Genexpression über eine dynamische Anpassung des tRNA-Modifikationssets als Ganzes denkbar. Die Suche nach solchen, das gesamte Modifikationsset betreffenden Dynamiken wird dadurch erschwert, dass dafür eine verlässliche Methode zur parallelen und exakten Quantifizierung einer größeren Zahl von tRNA-Modifikationen benötigt wird. Von Carell und Mitarbeitern wurde deshalb eine Quantifizierungsmethode entwickelt, die genau diesen Anforderungen entspricht.^[97] Es handelt sich hierbei um eine LC-MS-basierte Isotopenverdünnungsmethode, bei der die Zugabe von schweratommarkierten Referenznukleosiden eine genaue Quantifizierung der entsprechenden natürlichen Nukleoside in der Probe erlaubt (siehe Kapitel 4.1). Mit Hilfe dieser Methode konnte von Carell und Mitarbeitern gezeigt werden, dass *E. coli*^[98] und *S. cerevisiae*^[99] auf bestimmte Umweltreize mit einer Anpassung ihrer tRNA-Modifikationssets reagieren. Weiterhin wurde nachgewiesen, dass Säugetiere einen gewebespezifischen tRNA-Modifikationsgrad besitzen.^[100] Dieser ist umso höher, je größer die Proteinsyntheserate des entsprechenden Gewebes ist (siehe Kapitel 4.2). Wie die Anpassung des tRNA-Modifikationsgrades an bestimmte Umweltbedingungen bzw. an den Translationsbedarf eines Gewebes gesteuert wird, ist allerdings bisher noch völlig unbekannt.

3 Modifizierte DNA-Nukleoside: Vorkommen, biologische Bedeutung und Einfluss auf die Genexpression

3.1 Eine neue Familie DNA-modifizierender Enzyme

Lange Zeit war die Base 5-Methylcytosin (mC) die einzig bekannte enzymgenerierte DNA-Modifikation in Säugetieren.^[15] Die Entdeckung der enzymatischen Aktivität des Enzyms TET1,^[23, 101] welches durch Oxidation von mC unter anderem die DNA-Base 5-Hydroxymethylcytosin (hmC)^[22, 24] generiert, war der Ausgangspunkt für eine Vielzahl von bahnbrechenden Untersuchungen im Bereich der Epigenetik.^[49, 102] Dabei hat sich gezeigt, dass die Enzyme der TET-Familie^[29, 78, 103] und die ihnen generierten modifizierten DNA-Basen -neben den DNA-Methyltransferasen (DNMTs)^[104-105] und mC- eine Schlüsselrolle bei der epigenetischen Regulation der Genexpression spielen.^[49, 106] Sie beeinflussen nicht nur die Embryonalentwicklung,^[107-108] sondern sind z.B. auch notwendig für die Aufrechterhaltung der Pluripotenz von Stammzellen.^[109] Die bei der TET-katalysierten Oxidation von mC gebildeten Cytosinderivate besitzen u.a. eine herausragende Bedeutung als Intermediate der DNA-Demethylierung und können das zelluläre Transkriptionspotential beeinflussen.^[102]

3.1.1 Die TET-Enzyme: Substrate und Produkte, katalysierte Reaktionen und Struktur

Die Enzyme der TET-Familie^[23, 29, 101, 108, 110-111] (TET = *ten-eleven translocation*) sind Fe(II)- und α -Ketoglutarat (α -KG) abhängige Dioxygenasen.^[112] Als 5-Methylcytosinoxidasen sind sie verantwortlich für die schrittweise Oxidation der DNA-Base 5-Methylcytosin zu 5-Hydroxymethylcytosin,^[23] 5-Formylcytosin (fC)^[25-26] und 5-Carboxycytosin (caC)^[26-27] (siehe Abbildung 5). Das Hauptprodukt der TET-katalysierten Oxidation von mC ist Hydroxymethylcytosin, dessen Anteil im Gehirn der Maus, bezogen auf alle Cytosine und Cytosinmodifikationen, bis zu 0,8%,^[113] beim Menschen sogar bis zu 1,2%^[114-115] beträgt. Formylcytosin und besonders Carboxycytosin treten dagegen in deutlich geringeren Mengen auf, so dass nur etwa 0,02% bzw. 0,002% aller Cytosine formyliert bzw. -wenn überhaupt- carboxyliert sind.^[31] Vor kurzem konnten Carell und Mitarbeiter zeigen, dass die TET-Enzyme durch Oxidation von Thymin auch für die Bildung der DNA-Base 5-Hydroxymethyluracil (hmU) verantwortlich sind,^[31] welche zuvor nur als oxidativer DNA-Schaden bekannt war^[116-117] (siehe Abbildung 5).

TET-Enzyme kommen in allen vielzelligen Organismen (Metazoa) vor.^[101] Eine Ausnahme bilden einige Vielzeller wie z.B. *C. elegans*, bei denen die TET-Enzyme zusammen mit der

Cytosinmethylierung im Laufe der Evolution verlorengegangen sind.^[101] Bei den Kiefernmäulern (Gnathostomata) kam es zu einer Triplikation des Tet-Gens, so dass Säugetiere drei Tet-Enzyme (TET1-3) besitzen.^[101, 118]

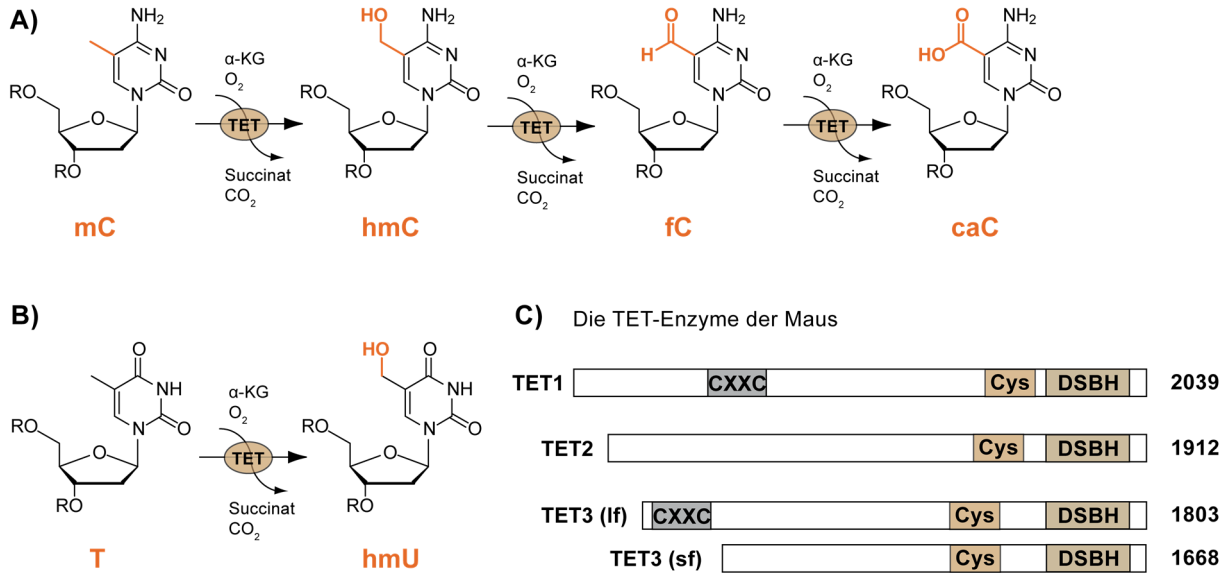


Abbildung 5. A) TET-katalysierte Oxidation von Methylcytosin (mC) zu Hydroxymethylcytosin (hmC), Formylcytosin (fC) und Carboxycytosin (caC); abgebildet ist jeweils das gesamte Nukleosid. B) TET-katalysierte Oxidation von Thymin (T) zu Hydroxymethyluracil (hmU); abgebildet ist jeweils das gesamte Nukleosid. C) Die TET-Enzyme der Maus mit farbiger Hervorhebung der wichtigsten Domänen.^[101] Neben einer längeren TET3-Variante mit CXXC-Domäne [TET3 (lf)] gibt es auch eine um 135 Aminosäuren kürzere Splicingform, welche die CXXC-Domäne nicht enthält [TET3 (sf)].^[119] α -KG = α -Ketoglutarat.

Die katalytische Domäne der TET-Enzyme befindet sich am C-terminalen Ende des Proteins (siehe Abbildung 5) und nimmt eine doppelsträngige β -Helixkonformation (DSBH) ein.^[101, 120-123] Nahe der DSBH-Domäne, welche typisch für Fe(II)- und α -KG-abhängige Dioxygenasen ist,^[112] befindet sich ein evolutionär konservierter cysteinreicher Abschnitt. Dieser besteht aus neun Cysteinresten und einem Histidinrest, welche zwei oder mehr Zn(II)-Ionen chelatisieren. Dieser Abschnitt ist an der Bindung der katalytischen Domäne zur DNA und an der Substraterkennung beteiligt. Die Bindung des Cofaktors α -Ketoglutarat erfolgt über eine Salzbrücke zu einem konservierten Argininrest. Das Fe(II)-Ion wird durch zwei Histidinreste und einen Aspartatrest komplexiert.^[101, 120-123] Ein weiteres Charakteristikum der TET-Enzyme ist ihre DNA-bindende Zinkfinger bzw. CXXC-Domäne.^[124] Sie besteht aus ca. 60 Aminosäuren und befindet sich bei TET1 und TET3 am N-Terminus der Peptidkette.^[120-122] Bei den Gnathostomata, zu denen auch die Säugetiere gehören, kam es im Laufe der Evolution zu einer chromosomalen Inversion im Bereich des Tet2-Gens.^[101] Dabei wurde das Exon mit der CXXC-Domäne von TET2 vom Rest des Gens getrennt und zu einem eigenständigen Gen, welches das mit TET2 interagierende CXXC4 bzw. IDAX-Protein kodiert.^[125] Interessanterweise gibt es in höheren

Säugetieren neben der die CXXC-Domäne enthaltenden TET3-Variante eine um 135 Aminosäuren kürzere Splicingform ohne CXXC-Domäne.^[119]

3.1.2 Mechanismus der TET-katalysierten Oxidation von Methylcytosin

Zu Beginn der TET-katalysierten Oxidation von mC (siehe Abbildung 6) wird das Substrat aus der DNA herausgedreht und in der katalytischen Tasche des Enzyms in unmittelbarer Nähe des Fe(II)-Ions positioniert. Der weitere Katalysemechanismus^[120-123] ist typisch für Fe(II)- und α -Ketoglutarat (α -KG)-abhängige Dioxygenasen.^[112, 126]

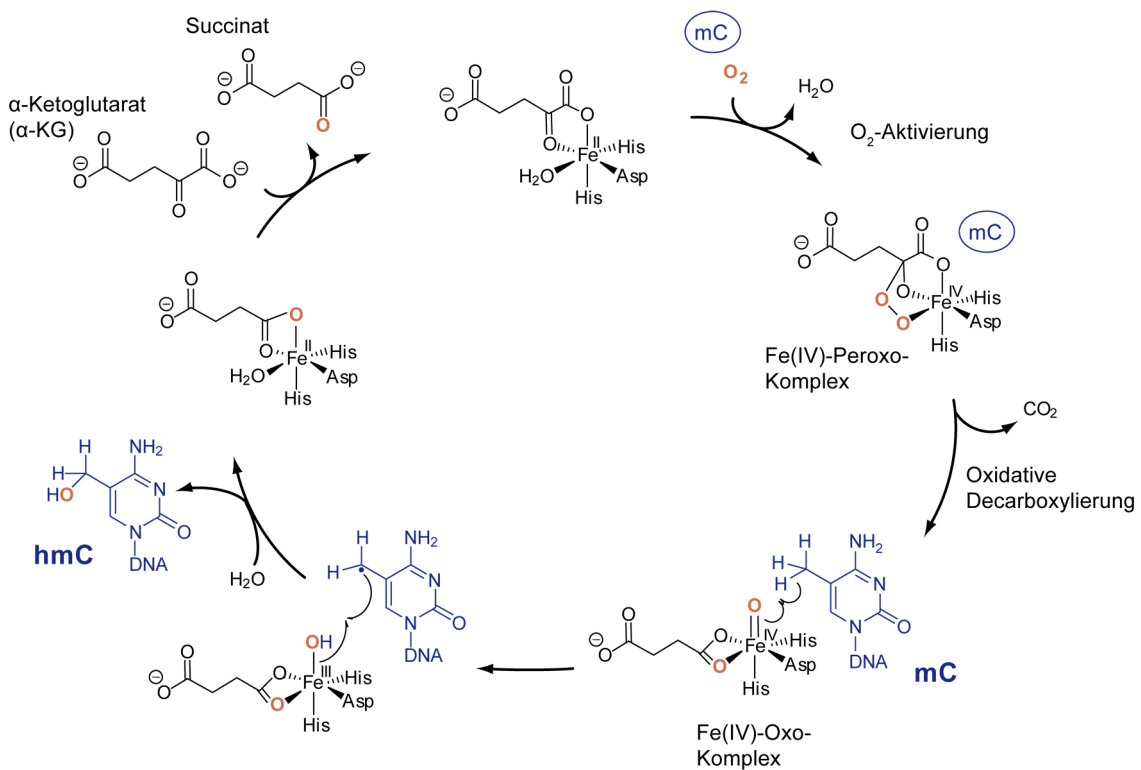


Abbildung 6. Katalysezyklus der TET-katalysierten Oxidation von 5-Methylcytosin (mC) zu 5-Hydroxymethylcytosin (hmC).

Zunächst erfolgt die Aktivierung von Sauerstoff am Fe(II)-Zentrum, bei der ein Fe(IV)-Peroxo-Komplex gebildet wird. Bei der anschließenden oxidativen Decarboxylierung des an das Fe(IV) gebundenen α -Ketoglutarats bildet sich ein Fe(IV)-Oxo-Komplex und CO₂ wird freigesetzt. Das gleichzeitig entstehende Succinat bleibt zunächst am Fe(IV)-Zentrum gebunden. Dann wird das Sauerstoffatom des Oxo-Komplexes radikalisch in zwei Schritten auf die Methylgruppe des mC übertragen, wobei als Zwischenstufe ein Fe(III)-Hydroxo-Komplex auftritt und das Fe(IV)-Zentrum zum Fe(II)-Zentrum reduziert wird. Das bei dieser Reaktion gebildete Hydroxymethylcytosin verlässt anschließend die

katalytische Tasche des Enzyms. Im finalen Schritt des Katalysezyklus kommt es zu einem Ligandenaustausch, wobei das zuvor gebildete Succinat durch ein neues Molekül α -Ketoglutarat ersetzt wird.^[120-123] Die Oxidation von hmC zu fC bzw. von fC zu caC oder von T zu hmU läuft nach einem analogen Mechanismus ab.

3.1.3 Nichtenzymatische biologische Wirkungen der TET-Enzyme

Interessanterweise ist die biologische Wirkung der TET-Enzyme nicht auf die Synthese der epigenetisch wirksamen Cytosinmodifikationen beschränkt. So können TET-Enzyme die Transkription offenbar auch nach einem Mechanismus beeinflussen, der unabhängig von ihrer enzymatischen Aktivität ist.^[29, 78, 102] Dazu passt die Beobachtung, dass die bevorzugten DNA-Bindungsstellen der TET-Enzyme nicht mit den mC- oder hmC-reichen DNA-Abschnitten zusammenfallen. Über ihre CXXC-Domäne binden alle drei TET-Enzyme bevorzugt an CpG-reiche Genpromotoren.^[127-128] Dies sind DNA-Abschnitte mit einer durchschnittlichen Länge von 1000 Basenpaaren, welche eine deutlich erhöhte CpG-Dinukleotiddichte besitzen und bei aktiv transkribierten Genen normalerweise hypomethyliert sind.^[19-20] Sie weisen außerdem eine geringe lokale hmC-Dichte auf.

Je nach TET-Typ hat diese Bindung unterschiedliche Auswirkungen auf die Expression des assoziierten Gens: Während es Hinweise darauf gibt, dass die Bindung von TET1 und TET3 an den Promotor zu einer verminderten Expression des entsprechenden Gens in embryonalen Stammzellen führt,^[127] verstärkt das Vorhandensein von TET2 in der Promotorregion die Expression des assoziierten Gens.^[129] Dieser gegensätzliche Einfluss auf die Genexpression ist zunächst überraschend. Er lässt sich jedoch dadurch erklären, dass TET1 und TET3 einerseits und TET2 andererseits mit unterschiedlichen Bindungspartnern interagieren, die wiederum die Transkription auf unterschiedliche Weise beeinflussen können. So rekrutiert TET1 den repressiven SIN3-Komplex zu den Genpromotoren.^[127, 130] SIN3 überführt das Chromatin mit Hilfe seiner Histondeacetylase-Untereinheiten in einen dichteren Packungszustand und fungiert so als lokaler Transkriptionsrepressor.^[131-134] TET1 interagiert außerdem mit dem NuRD-Komplex,^[130, 135] der genau wie SIN3 durch Chromatinmodifikation die Transkription inhibiert.^[136-138]

Die aktivierende Wirkung von TET2 auf die Genexpression wird u.a. über das OGT-Enzym^[139] vermittelt. OGT bindet an TET2 und überträgt den Zucker β -D-N-Acetylglucosamin^[140] auf Serin- oder Threoninreste verschiedener chromatinmodifizierender Proteine, was deren Aktivität beeinflusst.^[129, 141-142] Daneben kann OGT (neben den TET-Enzyme selbst,^[130, 143-145] siehe dazu Abschnitt 3.1.4.2) den Serinrest 112 von Histon H2B modifizieren.^[146] Diese Histonmodifikation ist mit Bereichen aktiver Transkription assoziiert.^[146] Es konnte gezeigt werden, dass ein Absenken des TET2-Levels zu einem

starken Rückgang der β -D-N-Acetylglucosaminmodifikation von H2BSer112 führt, was mit einer verminderten Genexpression einhergeht.^[146]

3.1.4 Expressionslevel der TET-Enzyme und Regulation ihrer Aktivität

3.1.4.1 Expressionslevel der TET-Enzyme in Organen und Zelltypen

Offensichtlich hat also jedes der drei TET-Enzyme eine spezifische biologische Funktion. Entsprechend variiert die Expression der einzelnen TET-Formen in den verschiedenen Entwicklungsstadien eines Organismus und in unterschiedlichen Organen bzw. Zelltypen.^[29, 78, 102] So ist TET3 besonders stark in der Zygote exprimiert^[147] und dort verantwortlich für die mC-Oxidation im paternalen Pronukleus.^[148-149] Diese leitet die Demethylierung der paternalen DNA ein, welche notwendig für die Wiederherstellung der Zellpluripotenz und bis zum 16-Zellstadium abgeschlossen ist (siehe auch Kapitel 3.4.1.1).^[150] TET1 und TET2 sind dagegen im Blastulastadium (Tag 5 der Embryonalentwicklung) stark exprimiert.^[106, 151] Zu diesem Zeitpunkt kommt es in der inneren Zellmasse, aus der sich später Embryo entwickelt, zu einer umfangreichen *de-novo*-Methylierung.^[152-153] Möglicherweise sind TET1 und TET2 hier an der Feinabstimmung des Methylierungsmusters beteiligt. Eine zweite globale Demethylierungswelle findet zwischen den Tagen 9,5 und 13,5 der Embryonalentwicklung bei der Bildung der Urkeimzellen statt (siehe auch Kapitel 3.4.1.1).^[154-157] In diesem Falle erfolgt die mC-Oxidation auch durch TET1 und TET2, welche im Gegensatz zu TET3 zu diesem Zeitpunkt in den Urkeimzellen stark exprimiert werden.^[108]

Im adulten Säugetier ist das TET-Expressions- und auch das hmC-Level im Gehirn^[18, 113] und hier wiederum in den Neuronen^[115] besonders groß. Es mehren sich die Hinweise darauf, dass die TET-Enzyme hier an lokusspezifischen DNA-Demethylierungen beteiligt sind.^[158-159] Diese sind die epigenetische Voraussetzung für neuronale Plastizität und ein normales Lernvermögen (siehe auch Kapitel 3.4.1.2).^[158-163] Wird z.B. TET1 im Hippocampus durch einen shRNA-vermittelten *in-vivo*-Knockdown gezielt herunterreguliert, so zeigen Mäuse eine deutlich verminderte Übertragung von Lerninhalten ins Langzeitgedächtnis.^[33]

Obwohl alle drei TET-Enzyme für sich genommen wichtige biologische Aufgaben erfüllen, ist interessanterweise nur ein Tet3-Doppelknockout (Tet3^{-/-}) letal. Mäuse mit einem solchen Tet3^{-/-} sterben aus bisher nicht geklärter Ursache um den Zeitpunkt der Geburt herum.^[147] Dagegen sind Mäuse mit Tet1^{-/-}, Tet2^{-/-} und sogar einem kombinierten Tet1- und Tet2-Doppelknockout^[164] lebensfähig. Allerdings weisen diese Mäuse z.T. starke Entwicklungsdefizite auf.

3.1.4.2 Regulation und Kontrolle der TET-Aktivität

Die Regulation der Expressionsrate ist eine von mehreren Möglichkeiten zur Kontrolle der TET-Aktivität.^[102] Tatsächlich gibt es auch Hinweise auf andere regulative Mechanismen, beispielsweise über die zelluläre Verfügbarkeit der Cofaktoren, hier besonders von α -Ketoglutarat.^[165] Außerdem konnte gezeigt werden, dass die Prozessivität der TET-Enzyme durch Vitamin C beeinflusst werden kann.^[166-167] So erhöhen sich bei Anwesenheit von Vitamin C im Kulturmedium von murinen embryonalen Stammzellen die Level aller mC-Oxidationsprodukte. Dieser Effekt betrifft besonders das Vorkommen von fC und caC, deren Level um den Faktor 5-10 ansteigt.

Daneben wird die Aktivität der TET-Enzyme offenbar auch durch posttranslationale Ubiquitylierung^[168], Phosphorylierung^[143] und O-GlcNAcylierung^[130, 143-145] gesteuert. Außerdem sind Proteine identifiziert worden, welche die TET-Enzyme kontextspezifisch zur DNA dirigieren können und damit ihre Aktivität steuern.^[169] So führt eine Rekrutierung von TET3 durch den Transkriptionsrepressor REST / NRSF^[170-174] zu einer Hydroxymethylierung der entsprechenden DNA-Abschnitte, welche mit einer Induktion der Genexpression einhergeht (siehe auch Kapitel 4.11).^[169]

3.1.5 TET-Mutanten, TET-Fehlregulation und damit einhergehende Erkrankungen

Es ist seit langem bekannt, dass eine fehlerhafte DNA-Methylierung zur Krebsentstehung beitragen kann.^[175] Ein anomales DNA-Methylierungsmuster (Methylom) gilt als ein wichtiges Kennzeichen einer malignen Gewebeveränderung.^[176-179] Mittlerweile gibt es Hinweise darauf, dass ein solches anomales Methylom auch durch eine Störung der TET-Funktion und damit der TET-vermittelten DNA-Demethylierung hervorgerufen werden kann.^[180-188] In diesen Fällen ist oft auch die DNA-Hydroxymethylierung betroffen, so dass Karzinome im Vergleich zu gesundem Gewebe sehr niedrige hmC-Level aufweisen (siehe auch Kapitel 4.8).^[114, 189-190]

Ursache für eine Störung der TET-Funktion kann z.B. eine Mutation in einem der drei Tet-Gene sein.^[191] So kann eine Mutation des humanen Tet2-Gens das myelodysplastische Syndrom auslösen.^[192] Dabei handelt es sich um eine Vorstufe der Leukämie, bei der immer mehr unreife Blutzellen gebildet werden.^[193-194] Entsprechend beobachtet man bei Mäusen mit Tet2-Knockout eine verstärkte Proliferation von hämatopoetischen (blutzellbildenden) Stammzellen.^[192]

Daneben kann die Funktionsstörung der TET-Enzyme auch durch einen Mangel des Cofaktors α -Ketoglutarat verursacht werden. α -Ketoglutarat wird durch die Enzyme Isocitrat-Dehydrogenase1 und 2 (IDH1/2) aus Isocitrat gebildet.^[195] Diese Reaktion ist Teil des Citratzyklus und damit ein zentraler Bestandteil des Energiestoffwechsels der Zelle.^[196-197] Weiterhin können auch die Idh-Gene von einer Mutation betroffen sein.^[195, 198] Die mutierten IDH-Enzyme können kein α -Ketoglutarat mehr bilden,

sondern wandeln das Isocitrat stattdessen zu 2-Hydroxyglutarat um.^[195] Der Onkometabolit 2-Hydroxyglutarat ist ein kompetitiver Inhibitor der TET-Enzyme und anderer α -Ketoglutarat-abhängiger Dioxygenasen.^[199-200] Zusammen mit dem Mangel an α -Ketoglutarat kommt es so zu einer Störung der normalen TET-Aktivität.^[186-187, 201] Das Resultat ist ein verringertes hmC-Level und ein anomales Methylom. Ein durch Mutation der IDH-Enzyme verursachtes anomales Methylom ist in vielen Fällen von akuter myeloischer Leukämie und daneben bei bestimmten Gliomen zu beobachten.^[191, 195, 198] Auch Mutationen anderer Schlüsselenzyme des Citratzyklus können die TET-Funktion negativ beeinflussen.^[202] So führen inaktivierende Mutationen der Enzyme Fumarat-Hydratase^[203-205] bzw. Succinat-Dehydrogenase,^[206] die man in bestimmten Krebszellen findet, zu einer Anhäufung von Fumarat bzw. Succinat in der Zelle. Sowohl Fumarat als auch Succinat inhibieren α -Ketoglutarat-abhängige Dioxygenasen und damit auch die TET-Enzyme.^[199]

3.2 Globale Quantifizierung und genomische Sequenzierung der Basen hmC, fC, caC und hmU

Um die biologische Bedeutung von hmC und der anderen neu entdeckten Cytosinmodifikationen zu verstehen, mussten zunächst grundlegende Kenntnisse über das Vorkommen dieser Basen in Organen und Geweben gewonnen werden. In diesem Kapitel werden die zu diesem Zwecke entwickelten Methoden und Techniken beschrieben.^[207-209] Es werden zunächst Methoden zur globalen Quantifizierung der Basen vorgestellt. Die Ergebnisse dieser Quantifizierungen bildeten die Grundlage für weitere Arbeiten zur Kartierung der Cytosinmodifikationen innerhalb des Genoms. Die dafür entwickelten Verfahren lassen sich in Methoden mit mittlerer Auflösung, welche das Prinzip der Affinitätsanreicherung nutzen, und in Sequenzierungsmethoden mit basengenaue Auflösung einteilen.

Die Ergebnisse der in diesem Kapitel vorgestellten Quantifizierungs- und Sequenzierungsmethoden, d.h. das Vorkommen der Cytosinmodifikationen in Organen und Geweben und ihre Verteilung im Genom, werden dann in Kapitel 3.3 besprochen.

3.2.1 Quantifizierung der globalen Cytosinmodifikationslevel in Organen und Geweben

Die ersten globalen Quantifizierungen von hmC basierten auf einer dünnschichtchromatographischen Methode und waren noch relativ ungenau.^[22-23] Eine exakte Bestimmung der globalen hmC- (und mC)-Level in Organen und Hirngewebe gelang erstmals mit Hilfe eines von Carell und Mitarbeitern entwickelten massenspektrometrie-basierten Verfahrens (siehe Abbildung 7).^[18, 113] Bei diesem Verfahren werden stabil isotoopenmarkierte Nukleosid-Standards, welche zuvor von Carell und Mitarbeitern synthetisiert werden mussten,^[210-211] zur Probe hinzugegeben. Anschließend wird das Probengemisch flüssigchromatographisch aufgetrennt und massenspektrometrisch analysiert. Die Nukleosid-Standards besitzen dasselbe chromatographische Verhalten und dieselben Ionisationseigenschaften wie die natürlichen Nukleoside, haben jedoch ein anderes Molekulargewicht. Damit wird eine separate Integration der Massensignale der natürlichen Nukleoside und der isotoopenmarkierten Standards möglich. Durch Vergleich dieser Integrale und unter Verwendung einer Eichgeraden lassen sich so die natürlichen Modifikationen in der Probe genau und parallel quantifizieren.

Nach demselben Prinzip konnten auch die absoluten Level der DNA-Modifikationen fC, caC und hmU in verschiedenen Organen und Geweben von Carell und Mitarbeitern exakt quantifiziert werden.^{[25, 31,}

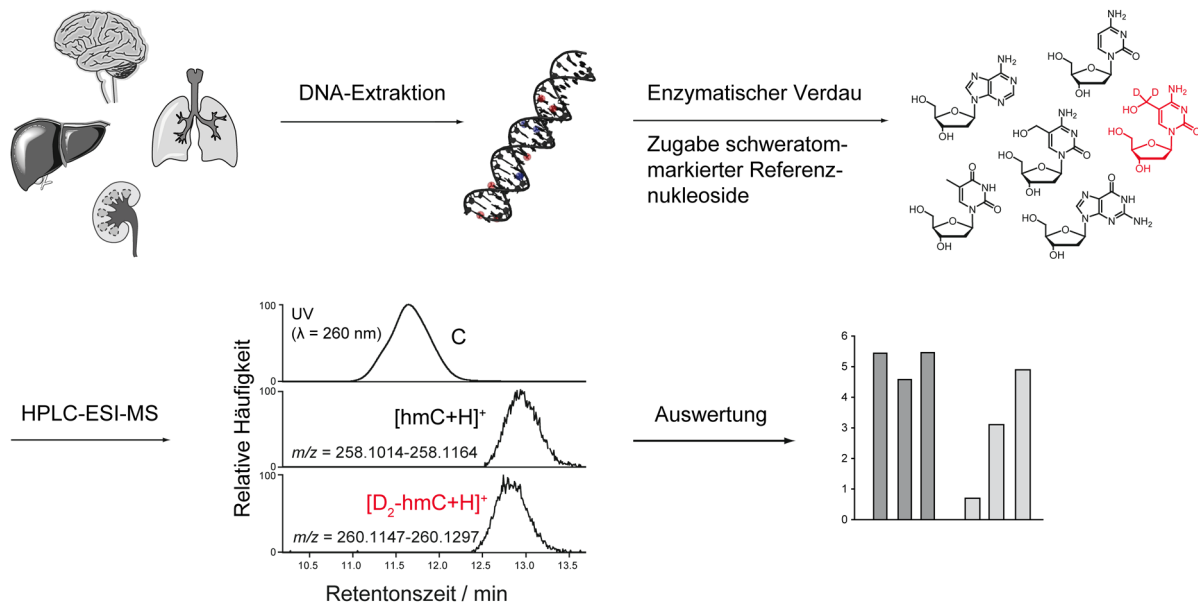


Abbildung 7. Schematische Darstellung des von Carell und Mitarbeitern entwickelten Verfahrens zur exakten und quantitativen Bestimmung des globalen hmC-Llevels in biologischen Proben.^[18, 113]

Eine etwas später publizierte Methode zur globalen Quantifizierung von hmC beruht auf der enzymatischen Übertragung von radioaktiv markierter Glucose auf die OH-Gruppe der hmC-Basen.^[212] Im Vergleich zu der von Carell und Mitarbeitern entwickelten Methode erwies sich dieses Verfahren jedoch als weniger exakt.

3.2.2 Affinitätsanreicherungsmethoden

Bei den Affinitätsanreicherungsmethoden^[213] wird die genomische DNA zunächst in kürzere Abschnitte gespalten. Dann werden die Abschnitte, welche die gesuchte Cytosinmodifikation enthalten, von der Gesamt-DNA abgetrennt. Die Abtrennung erfolgt entweder durch Immunpräzipitation, wofür jedoch eine hohe Modifikationsdichte auf der DNA notwendig ist, oder unter Ausnutzung einer Biotin-Streptavidin-Wechselwirkung. Anschließend werden die auf diese Weise angereicherten DNA-Abschnitte amplifiziert und sequenziert. Das Ergebnis ist eine niedrig aufgelöste Karte der gesuchten Modifikation im Genom. Die niedrige Auflösung folgt aus der Länge der präzipitierten DNA-Abschnitte, die einige Hundert Basenpaare beträgt. Dafür handelt es sich jedoch um effiziente, mit relativ geringen Kosten verbundene Methoden.^[49, 214]

3.2.2.1 Methoden zur Affinitätsanreicherung von hmC

Unter dem Namen MeDIP (*methylated DNA immunoprecipitation*) wird das Prinzip der Immunpräzipitation routinemäßig für die Analyse von methylierten DNA-Fragmenten genutzt.^[215-216] Mittlerweile gibt es auch Antikörper für hmC, fC und caC, die jedoch z.T. unerwünschte unspezifische Wechselwirkungen eingehen können und jeweils nur an DNA-Abschnitte mit hoher Modifikationsdichte binden.^[127, 217-220] Spezifischer scheint dagegen ein Antikörper gegen Cytosin-5-methylsulfonat zu sein, welches bei der Reaktion von hmC mit Bisulfit entsteht.^[192, 221-222]

Bei einer weiteren Methode zur Affinitätsanreicherung von hmC-haltigen DNA-Abschnitten wird die Wechselwirkung zwischen glycosyliertem hmC (ghmC) und dem Protein JBP1 ausgenutzt.^[223] JBP1 ist ein Protein, welches in *Kinetoplastea* wie dem parasitären Eukarioten *Trypanosoma brucei* natürlich vorkommt und dort an die Base J (D-Glucopyranosyloxymethyluracil, ghmU) bindet.^[224-226] Das genomische hmC muss bei dieser Methode zunächst glycosyliert werden. Zur Glycosylierung wird dabei das Enzym β -Glycosyltransferase (β -GT) verwendet, das aus T4-Bakteriophagen isoliert wurde, in denen glycosyliertes hmC natürlicherweise vorkommt.^[212, 227]

Die oben vorgestellten Methoden zur Immunpräzipitation von DNA-Abschnitten mit Cytosinmodifikationen haben den Nachteil, dass z.T. unspezifische Wechselwirkungen auftreten können. Auch die bei der Verwendung von JBP1 erreichte Spezifität wird teilweise durch eine geringe Signalintensität aufgewogen, die aus der relativ schwachen Bindung zwischen JBP1 und ghmC resultiert.^[214] Mehrere Gruppen haben deshalb Sequenziermethoden für hmC entwickelt, welche die starke und spezifische Wechselwirkung zwischen Biotin und Streptavidin für die Affinitätsanreicherung ausnutzen. Um hmC zu biotinylieren, ist bei den hMe-Seal und GLIB genannten Methoden zunächst eine Glycosylierung der Hydroxylgruppe von hmC erforderlich.^[221, 228] Diese Glycosylierung erfolgt enzymatisch und wieder durch die Glycosyltransferase β -GT.^[212] Bei der hMe-Seal-Methode enthält die dabei übertragene Glucose bereits eine Azidogruppe, so dass die Base durch eine Klickreaktion mit Biotin modifiziert werden kann.^[228] Bei der GLIB-Methode wird die Glucose dagegen erst nach der Übertragung auf hmC funktionalisiert. Dazu wird die vicinale OH-Gruppe der Glucose mit Natriumperiodat zur Aldehydgruppe oxidiert. Dies erlaubt dann die Biotinylierung der Base durch Aminooxy-Biotin^[221] Hierbei kann es jedoch zu Fehlern kommen, da auch fC-Basen mit ihrer Aldehydgruppe oder abasische Stellen in der DNA (*AP-Sites* = *apurinic* bzw. *apyrimidinic sites*) nach einer Ringöffnung der Desoxyribose mit Aminooxy-Biotin reagieren können.

3.2.2.2 Methoden zur Affinitätsanreicherung von fC und hmU

Die Wechselwirkung zwischen Biotin und Streptavidin kann auch zur Anreicherung von fC-haltigen DNA-Abschnitten und damit für eine Sequenzierung von fC genutzt werden.^[229] Da fC eine Aldehydgruppe enthält, ist eine direkte Biotinylierung der Base durch Umsetzung der DNA mit Aminoxy-Biotin möglich. Dieser Umstand wurde beispielsweise bei der Entdeckung von fC durch Carell *et al.* dazu genutzt, den massenspektrometrischen Nachweis von fC zu verbessern.^[25] Allerdings kann Aminoxy-Biotin, wie oben erwähnt, auch mit abasischen Stellen innerhalb der DNA reagieren. Diese unerwünschte Nebenreaktion wird bei der fC-Seal genannten Sequenziermethode vermieden, was jedoch mit einem erheblichen Mehraufwand verbunden ist.^[230] Bei fC-Seal wird das gesamte genomische fC zu hmC reduziert, so dass die Base nach einer für hmC entwickelten und etablierten Methode sequenziert werden kann. Die fC-Seal-Methode erfordert daher zunächst ein *Capping* der endogenen hmC-Basen durch Glycosylierung, das enzymatisch mit Hilfe von β -GT durchgeführt wird. Anschließend wird das fC mit NaBH₄ zum hmC reduziert.^[230] Die so vorbehandelte DNA kann dann nach der hMe-Seal-Methode (siehe oben) sequenziert werden.^[228] Sollte allerdings die fC-Reduktion nicht quantitativ verlaufen, so führt die fC-Seal-Methode unweigerlich zu Fehlern.

In einer *Proof-of-principle*-Studie wurde von Hardisty *et al.* eine Methode zur Affinitätsanreicherung von 5-Formyluracil (fU) entwickelt.^[231] Diese Methode lässt sich prinzipiell auch für die hmU-Anreicherung nutzen, da sich Hydroxymethyluracil mit K₂Cr₂O₇ zu fU oxidieren lässt.^[231] Zuvor ist natürlich ein *Capping* des natürlich in der DNA vorkommenden fUs erforderlich. Im nächsten Schritt ermöglicht die höhere Reaktivität der Aldehydgruppe von fU im Vergleich zur Aldehydgruppe von fC eine chemoselektive Reaktion zwischen den fU-Basen und einem biotinylierten Linker, z.B. unter Oximbildung. Oligonukleotide mit den so markierten Basen lassen sich dann über eine Biotin-Streptavidin-Wechselwirkung anreichern und anschließend sequenzieren.

Wie Carell und Mitarbeiter gezeigt haben, wird das Gros der hmU-Basen zumindest in embryonalen Stammzellen durch TET-vermittelte Oxidation aus Thymin gebildet, so dass die meisten hmU-Basen hier im hmU:A-Kontext vorkommen. Die Desaminierung von hmC zu hmU, welche zu einer hmU:G-Fehlpaarung in der DNA führt, spielt dagegen nur eine geringe Rolle. Eine Methode zur Affinitätsanreicherung solcher fehlgepaarter hmU-Basen wurde kürzlich von He und Mitarbeitern vorgestellt.^[232]

3.2.3 Sequenziermethoden mit basengenaue Auflösung

Die Sequenziermethoden mit basengenaue Auflösung^[213] beruhen auf dem gleichen Prinzip wie die Bisulfitsequenzierung (BS-Seq), welche für die Kartierung des Methyloms entwickelt wurde (siehe Abbildung 8).^[233] Beim ursprünglichen BS-Seq wird die DNA zunächst in zwei Proben aufgeteilt, von denen eine mit Natriumbisulfit (Natriumhydrogensulfit) behandelt wird. In dieser Probe werden durch das Bisulfit alle in der DNA vorkommenden Cytosine zu Uracil desaminiert. mC wie auch A, T und G überstehen die Behandlung unverändert.^[234-235] Anschließend werden beide Proben *via* PCR amplifiziert, dabei erkennt die Polymerase wegen des Watson-Crick-Basenpaarungsverhaltens U als T und mC als C. Vergleicht man dann das Ergebnis der DNA-Sequenzierung beider Proben, so kann man aus den Einzelnukleotid-Polymorphismen auf die Art und Position der modifizierten Cytosine schließen. Beispielsweise muss es sich bei denjenigen Basen um mC handeln, welche in beiden Fällen als C gelesen wurden. Cytosin wird dagegen in der Bisulfitprobe als T und in der unbehandelten Referenzprobe als C gelesen.^[233] Liegt bereits ein Referenzgenom zum Vergleich vor, ist die Sequenzierung der unbehandelten Probe für die Kartierung des Methyloms natürlich nicht erforderlich.

Sequenzierung von mC		C	mC	hmC	fC	caC
PCR-Amplifikation der unbehandelten DNA		C	C	C	C	C
BS-Seq.	a) Behandlung der Probe mit Natriumbisulfit	U	mC	CMS	U	U
	b) PCR-Amplifikation	T	C	C	T	T

Abbildung 8. Sequenzierung von Methylcytosin nach der Bisulfit-Methode. Durch eine Sequenzierung der PCR-amplifizierten DNA von behandelter und unbehandelter Probe lässt sich die Position der mC-Basen (inklusive hmC) ermitteln. Orange: Ursprünglich in der DNA-Probe vorhandene Basen; braun und fettgedruckt: Ergebnis der Sequenzierung nach PCR-Amplifikation der DNA. PCR = Polymerase-Kettenreaktion; CMS = Cytosin-5-methylsulfonat; U = Uracil.^[233, 236-238]

Nach der Entdeckung von hmC zeigte sich, dass diese Cytosinmodifikation mit Bisulfit zu Cytosin-5-methylsulfonat (CMS) reagiert (siehe Abbildung 8). Dieses verhält sich bei einer Bisulfitsequenzierung genau wie mC.^[236-237] Dagegen werden fC und caC zunächst deformyliert bzw. decarboxyliert und dann, wie auch Cytosin, zu U desaminiert.^[238] Die mC-Datensätze müssen also mit Vorsicht betrachtet werden, da es sich bei den als methyliert ausgewiesenen Cytosinen auch um hmC handeln kann. Es wurden daher für die neuentdeckten Cytosinmodifikationen spezielle Sequenziermethoden mit basengenaue Auflösung entwickelt.^[49]

3.2.3.1 Basengenaue Sequenzierung von hmC

Eine Methode zur Sequenzierung von hmC nennt sich oxBS-Seq.^[239] Hier wird mit Hilfe von KRuO_4 das gesamte genomische hmC zunächst zu fC oxidiert. Bei der anschließenden Bisulfitbehandlung ist nur mC inert gegenüber einer Desaminierung. Vergleicht man das Ergebnis von PCR-Amplifikation und darauffolgender Sequenzierung mit dem Ergebnis einer normalen BS-Seq derselben Probe, so lässt sich mC direkt als C identifizieren. hmC dagegen wird beim BS-Seq als C und nach KRuO_4 -Behandlung als T gelesen (siehe Abbildung 9). Der große Nachteil dieser Methode ist ihr Materialverbrauch, da nicht nur bei der Bisulfitbehandlung, sondern auch bei der Oxidation durch KRuO_4 jeweils ein Großteil der DNA zerstört wird. Es muss außerdem immer sichergestellt werden, dass die hmC-Oxidation quantitativ erfolgt und keine Weiteroxidation zu caC auftritt.

Sequenzierung von hmC		C	mC	hmC	fC	caC
BS-Seq.	PCR-Amplifikation der unbehandelten DNA	C	C	C	C	C
	PCR-Amplifikation nach BS-Behandlung	T	C	C	T	T
oxBS-Seq.	a) Behandlung der Probe mit KRuO_4	C	mC	fC	fC	caC
	b) Behandlung der Probe mit Natriumbisulfit	U	mC	U	U	U
	c) PCR-Amplifikation	T	C	T	T	T
TAB-Seq.	a) Capping via Beta-GT	C	mC	ghmC	fC	caC
	b) TET1cd-vermittelte Oxidation	C	caC	ghmC	caC	caC
	c) Behandlung der Probe mit Natriumbisulfit	U	U	ghmC	U	U
	d) PCR-Amplifikation	T	T	C	T	T

Abbildung 9. Methoden zur Sequenzierung von Hydroxymethylcytosin.^[239-240] Durch eine Sequenzierung der PCR-amplifizierten DNA von behandelter, Bisulfit-behandelter und unbehandelter Probe lässt sich die Position der hmC-Basen ermitteln. Orange: Ursprünglich in der DNA-Probe vorhandene Basen; braun und fettgedruckt: Ergebnis der Sequenzierung nach PCR-Amplifikation der DNA. PCR = Polymerase-Kettenreaktion; BS = Bisulfit; U = Uracil; Beta-GT = β -Glycosyltransferase; ghmC = glycosyliertes hmC.

Bei der TAB-Seq-Methode (*Tet-assisted bisulfite sequencing*) zur Sequenzierung von hmC wird auch ein Oxidationsschritt durchgeführt.^[240] Dieser erfolgt enzymatisch mit Hilfe der katalytischen Domäne von TET1 (TET1cd). Dabei werden mC und fC vollständig zu caC umgesetzt, hmC ist dagegen als 5-(β -D-Glucopyranosyloxymethyl)cytosin (ghmC) vor einer Oxidation geschützt (siehe Abbildung 9). Dazu wird zu Beginn der Probenbehandlung mit Hilfe der Glycosyltransferase β -GT (siehe Abschnitt 3.2.2.1) Glucose als Schutzgruppe auf die Hydroxylgruppe von hmC übertragen. Zuletzt wird

eine Bisulfitsequenzierung der vorbehandelten Probe durchgeführt. hmC kann dabei direkt als C ausgelesen werden, da ghmC nicht durch Bisulfit desaminiert wird. Bei der TAB-Seq-Methode ist es von entscheidender Bedeutung, dass die enzymatische Oxidation quantitativ verläuft, so dass mC vollständig umgesetzt wird. Ansonsten wird auch mC (bzw. zu hmC oxidiertes mC) später als C ausgelesen und damit als hmC interpretiert. Dies mag besonders bei einer unsachgemäßen Handhabung des Enzyms vorkommen. Außerdem muss das Capping von hmC vollständig sein.

3.2.3.2 Basengenaue Sequenzierung von fC und caC

Der Schlüsselschritt der redBS-Sequenzierung (*reduced bisulfite sequencing*), einer Methode zur genomischen Kartierung von fC, ist die Reduktion des genomischen fC zu hmC.^[241] Als Reduktionsmittel wird Natriumborhydrid (NaBH₄) verwendet, welches caC nicht reduzieren kann. Mit der so behandelten Probe wird eine Bisulfitsequenzierung durchgeführt. Die genaue Position der fC-Basen ergibt sich dann aus dem Vergleich mit dem Ergebnis einer Standard-Bisulfitsequenzierung (siehe Abbildung 10).

Sequenzierung von fC oder caC		C	mC	hmC	fC	caC
BS-Seq.	PCR-Amplifikation der unbehandelten DNA	C	C	C	C	C
	PCR-Amplifikation nach BS-Behandlung	T	C	C	T	T
fC: redBS-Seq.	a) Behandlung der Probe mit NaBH ₄	C	mC	hmC	hmC	caC
	b) Behandlung der Probe mit Natriumbisulfit	U	mC	CMS	CMS	U
	d) PCR-Amplifikation	T	C	C	C	T
fC: fCAB	a) Behandlung der Probe mit EtO-NH ₂	C	mC	hmC	5-(RO-N=CH)C	caC
	b) Behandlung der Probe mit Natriumbisulfit	U	mC	CMS	5-(RO-N=CH)C	U
	d) PCR-Amplifikation	T	C	C	C	T
caC: caCAB	a) Behandlung der Probe mit EDC	C	mC	hmC	fC	"EDC+caC"
	b) Behandlung der Probe mit Natriumbisulfit	U	mC	CMS	U	"EDC+caC"
	d) PCR-Amplifikation	T	C	C	T	C

Abbildung 10. Methoden zur Sequenzierung von Formyl-^[230, 241] oder Carboxycytosin.^[242] Durch eine Sequenzierung der PCR-amplifizierten DNA von behandelter, Bisulfit-behandelter und unbehandelter Probe lässt sich die Position der fC- bzw. caC-Basen ermitteln. Orange: Ursprünglich in der DNA-Probe vorhandene Basen; braun und fettgedruckt: Ergebnis der Sequenzierung nach PCR-Amplifikation der DNA. PCR = Polymerase-Kettenreaktion; BS = Bisulfit; CMS = Cytosin-5-methylsulfonat; U = Uracil; EDC = 1-Ethyl-3-[3-dimethyl-aminopropyl]-carbodiimidhydrochlorid.

Die fCAB-Sequenzierungsmethode für fC und die caCAB-Methode zur Sequenzierung von caC enthalten beide einen *Capping*-Schritt, um die Desaminierung von fC bzw. caC bei der Bisulfitbehandlung zu verhindern (siehe Abbildung 10).^[230, 242] Bei der fCAB-Methode wird das fC durch Oximbildung mit einem Hydroxylamin (EtO-NH₂) geschützt.^[230] Dadurch kommt es im Rahmen der Bisulfitbehandlung nur bei caC zu einer Desaminierung. Nach PCR und Sequenzierung wird daher nur caC als T ausgelesen, während fC zusammen mit mC und hmC als C gelesen wird. Durch einen Vergleich der fCAB-Sequenz mit der Sequenz der Standard-Bisulfitsequenzierung können so die formylierten Cytosine ermittelt werden.

Die caCAB-Seq-Methode zur Sequenzierung von caC beginnt mit dem Capping von caC durch das Carbodiimid EDC.^[242] EDC (1-Ethyl-3-[3-dimethylaminopropyl]-carbodiimidhydrochlorid) bildet bei der Reaktion mit der Carboxylgruppe ein Amid. Das so geschützte caC wird durch Bisulfit nicht desaminiert. Auch beim caCAB-Seq werden die Positionen der caC-Basen durch Vergleich mit der Standard-Bisulfitsequenz bestimmt.

3.2.4 Sequenzierungsmethoden der dritten Generation

Vor einer Sequenzierung ist es bei allen zurzeit gängigen Methoden notwendig, die DNA *via* PCR zu amplifizieren. Dabei geht die über die modifizierten Cytosine kodierte epigenetische Information komplett verloren. Mit der Bisulfitmethode und ähnlichen Verfahren ist dennoch eine basengenaue Sequenzierung der Cytosinmodifikationen möglich. Diese erfordert jedoch eine chemische Vorbehandlung der Probe und ist daher umständlich und materialintensiv.^[233] Bei den sogenannten Sequenzierungsmethoden der dritten Generation^[213] entfällt dagegen der DNA-Amplifikationsschritt, da individuelle Moleküle sequenziert werden.^[243] Diese Methoden bieten sich besonders für eine Sequenzierung der modifizierten Cytosinbasen an.^[244]

3.2.4.1 SMRT-Sequencing

Die meisten Sequenzierungsmethoden der dritten Generation befinden sich zurzeit noch im Entwicklungsstadium. Ein bereits kommerzialisiertes Verfahren ist das SMRT-Sequencing (*Single-molecule real-time*; SMRT-Seq.) von *Pacific Biosciences*.^[245] Mit dieser Methode ist eine Sequenzierung sowohl von mC und hmC,^[246] als auch von fC und caC möglich.^[247] Beim SMRT-Verfahren wird ein DNA-Einzelstrang sequenziert, indem die Synthese seines Gegenstrangs durch eine Polymerase in Echtzeit detektiert wird. Die kanonischen Basen werden bei dieser Methode anhand ihrer charakteristischen Fluoreszenzmarkierung beim Einbau in die DNA erkannt.^[248] Eine

modifizierte Cytosinbase im Templatstrang wird dagegen dadurch identifiziert, dass die Polymerase an dieser Stelle für Synthese des Gegenstranges mehr Zeit als üblich benötigt. Die modifizierten Cytosine können daher umso besser identifiziert werden, je stärker ihr Einfluss auf die Kinetik der Polymerase ist.^[246] Eine zusätzliche chemische Modifizierung kann diesen Einfluss vergrößern und erleichtert so die Detektion dieser Basen nach der SMRT-Methode.^[247] Die Sequenzierung von mC kann beispielsweise über den Umweg einer Tet-katalysierten Oxidation der Base zu caC verbessert werden.^[249] Auch bei hmC liefert die Sequenzierung nach dem SMRT-Prinzip erst nach einer Modifizierung der Hydroxylgruppe mit einem Glukosederivat passable Ergebnisse.^[250]

Das SMRT-Sequencing eignet sich ebenso zur Sequenzierung von hmU. So konnten Korlach und Mitarbeiter zeigen, dass auch hmU-Basen im Templatstrang während der Synthese des Gegenstranges eine charakteristische kinetische Signatur erzeugen.^[251] Diese individuelle Signatur lässt sich vom Einfluss auf die Kinetik der Polymerase unterscheiden, welcher von anderen natürlichen DNA-Modifikationen oder oxidativ geschädigten DNA-Basen hervorgerufen wird. Das SMRT-Verfahren wurde darüber hinaus auch schon für eine Sequenzierung der mRNA-Modifikation m6A (siehe Kapitel 2.1) angewendet.^[252]

3.2.4.2 Nanopore-Sequencing

Neben dem kommerzialisierten SMRT-Seq.-Verfahren gibt auch es vielversprechende *Proof-of-Principle*-Studien zu anderen Sequenzierungsmethoden der dritten Generation.^[244] Bei Sequenzieretechnologien auf Grundlage von Nanoporen (*Nanopore-Sequencing*) beispielsweise wird der DNA-Strang beim Durchgang durch eine Membranpore sequenziert.^[253] Das detektierte Signal ist der Ionenstrom, welcher beim Anlegen einer Spannung durch die Pore fließt. Der DNA-Strang bzw. dessen Basen blockieren beim Durchgang durch die Pore einen Teil dieses Ionenstroms. Da die Abnahme des Ionenstroms u.a. von der Größe und Form der jeweiligen Base abhängt, kann aus den detektierten Intensitätsschwankungen des Ionenstroms direkt die Basensequenz ermittelt werden.^[254-256] Auf diese Weise lassen sich die Basen mC, hmC und C in verschiedenen experimentellen Systemen zuverlässig voneinander unterscheiden.^[257-259] Im bisher elaboriertesten dieser Systeme wird ein DNA-Einzelstrang von einer phi 29 DNA-Polymerase durch das mutierte bakterielle Porenprotein MspA gezogen. Durch Messung des Ionenstroms kann hierbei sowohl die genaue Position als auch die Art der modifizierten Cytosinbase (mC oder hmC) bestimmt werden.^[260]

Viele der bisher entwickelten Methoden zur Affinitätsanreicherung und zur basengenauen Sequenzierung der Cytosinderivate und von hmU erfordern eine chemische Modifikation dieser

Basen.^[213] Sollte der Modifikationsschritt nicht quantitativ verlaufen –was oft der Fall ist–, so führen diese Methoden zu Fehlern. Es werden deshalb trotz der Vielzahl an verfügbaren Methoden immer wieder neue Affinitätsanreicherungs-^[232, 247, 261-262] und Sequenzierungsverfahren^[208, 244, 263-267] vorgestellt, welche im besten Falle bessere Resultate liefern, als bereits bekannte Methoden.

3.3 Vorkommen und Verteilung von hmC, fC, caC und hmU im Genom

Die im obigen Kapitel beschriebenen Quantifizierungs- und Sequenzierungsmethoden haben unser Wissen über die Cytosinmodifikationen hmC, fC und caC enorm erweitert. Die mit Hilfe dieser Verfahren gewonnenen Kenntnisse über das Vorkommen dieser Basen in Organen und Geweben und ihre Verteilung im Genom werden in diesem Kapitel dargestellt.

3.3.1 5-Hydroxymethylcytosin (hmC)

Die erste genaue und absolute Bestimmung globaler hmC-Level wurde von Carell und Mitarbeitern in Hirngeweben^[113] und Organen^[18] von Mäusen durchgeführt. Dabei zeigte sich, dass hmC in allen untersuchten Organen vorkommt und dass das hmC-Level zwischen den Organen stark variiert. Während in den meisten Organen der Maus nur etwa 0,05-0,15% aller Cytosine hydroxymethyliert sind, wurden im Gehirn der Maus hmC-Level von 0,35-0,75% detektiert. Der Anteil von mC, bezogen auf alle Cytosine und Cytosinmodifikationen, liegt dagegen konstant bei ca. 4,5%.^[18, 31]

Carell und Mitarbeiter haben außerdem als Erste einen Anstieg des hmC-Levels in Abhängigkeit vom Lebensalter beobachtet.^[113] Mit zunehmendem Alter steigt die hmC-Modifikationsdichte in den Organen und Geweben zunächst stetig an, um dann im voll entwickelten Organismus ein konstant hohes Niveau zu erreichen.^[115, 228, 268-277] Im cerebralen Cortex sind dann beim Menschen etwa 1,2% und bei der Maus 0,6-0,8% aller Cytosine hydroxymethyliert^[115] (siehe Abbildung 11; eine detaillierte Beschreibung des postnatalen Anstiegs der hmC-Level in Säugetieren findet sich in den Kapiteln 4.6 und 4.7).

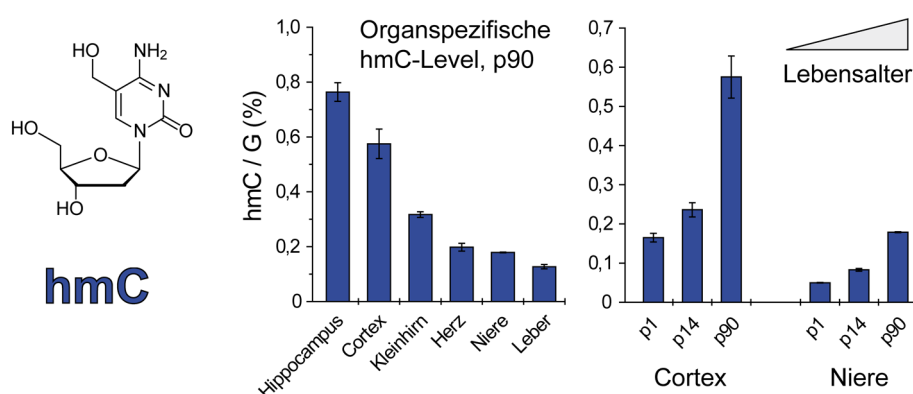


Abbildung 11. Organspezifische Level^[31] und altersabhängige Zunahme^[115] von hmC in der Maus. Alle Werte sind als *Modifikationen pro hundert Guaninbasen* angegeben. Guanin (G) wurde als Referenz gewählt, weil die Zahl der Guaninbasen der Summe aus Cytosin (C) und den Cytosinderivaten mC, hmC und fC entspricht. p1 = postnataler Tag 1, d.h. erster Tag nach der Geburt, p14 und p90 entsprechend.

Einen Überblick über die Verteilung von hmC im humanen und murinen Genom konnte mit Hilfe der in Kapitel 3.2 beschriebenen Sequenziermethoden gewonnen werden. Dabei ist zu beachten, dass diese hmC-Sequenzierungen immer nur *Momentaufnahmen* der Verteilung von hmC im Genom liefern. Möglicherweise auftretende dynamische Hydroxymethylierungen können mit diesen Methoden nicht von dauerhaft hydroxymethylierten Cytosinen unterschieden werden.

Untersucht wurden zunächst verschiedene Hirngewebe und embryonale Stammzellen, in denen hmC jeweils in großen Mengen vorkommt.^[31, 113] Die verschiedenen Sequenziermethoden lieferten dabei Großteils übereinstimmende Ergebnisse und es zeigte sich, dass das hmC im Wesentlichen in den euchromatischen Regionen des Genoms lokalisiert ist.^[47, 278] Dies sind Bereiche mit hoher Gendichte und Transkriptionsaktivität, in denen die DNA weniger dicht gepackt ist.^[279-280] Innerhalb der euchromatischen Bereiche findet man hmC angereichert in distalen regulatorischen Elementen, in Genpromotoren, in den Genkörpern und an Transkriptionsstartpunkten.^[49, 278]

Nahezu die Hälfte des gesamten genomischen hmCs ist mit sogenannten distalen *cis*-regulatorischen Elementen assoziiert.^[240, 281-283] Bei diesen Elementen handelt es sich um distal zum regulierten Gen gelegene Sequenzabschnitte mit Bindungsstellen für Transkriptionsfaktoren.^[240, 284] Eine basengenaue Sequenzierung dieser Bereiche zeigte, dass zwar die Cytosine in direkter Nähe zu den Transkriptionsfaktorbindungsstellen größtenteils hydroxymethyliert sind, die Bindungsstellen selber jedoch praktisch kein hmC enthalten.^[240, 282]

Außer in distalen regulatorischen Elementen und Promotoren ist hmC auch an Transkriptionsstartpunkten und in den Genkörpern zu finden.^[47, 278] In allen untersuchten Zell- und Gewebetypen ist das Vorhandensein von hmC im Genkörper dabei immer mit einer aktiven Transkription des Gens verbunden.^[127, 217-218, 228, 282-283, 285] Das hmC befindet sich hier eher in der an das 3'-Ende angrenzenden Hälfte des Gens und ist in murinen embryonalen Stammzellen besonders in den Exons angereichert.^[219, 221, 272]

Hydroxymethylierte Cytosine treten praktisch ausschließlich (>99%) als Bestandteil von CpG-Dinukleotiden (CpGs) auf.^[240] Auch Cytosinmethylierungen finden besonders häufig an CpGs statt,^[286] jedoch kann mC im Gegensatz zu hmC auch außerhalb des CpG-Kontextes vorkommen. So findet man in humanen embryonalen Stammzellen bis zu 25 % aller methylierten Cytosine in der Sequenz ^mCHG oder ^mCHH (H = A, C oder T).^[282, 287] Methylcytosine treten im DNA-Doppelstrang in den palindromischen CpG-Dinukleotiden immer symmetrisch auf, d.h. wenn ein bestimmtes CpG-Dinukleotid methyliert ist, dann ist das GpC-Dinukleotid im Gegenstrang auch methyliert.^[20] Eine Hydroxymethylierung der CpG-Sequenzpalindrome ist dagegen zumindest in mESCs z.T. asymmetrisch.^[264] In diesen Fällen findet man das hmC eher im Einzelstrang mit der höheren lokalen Guanindichte.^[240] Eine basengenaue hmC-Sequenzierung des mESC-Genoms ergab, dass das

gleiche Cytosin innerhalb einer Zellpopulation oft sowohl hydroxymethyliert als auch methyliert vorkommen kann.^[240] Generell sind potentielle Hydroxymethylierungspositionen durchschnittlich nur zu etwa 20% hydroxymethyliert, während potentielle Methylierungspositionen zu etwa 60% methyliert sind.^[240]

Vermutlich wegen der Mutagenität der Desaminierung von mC zu T sind CpG-Dinukleotide in den meisten DNA-Abschnitten statistisch unterrepräsentiert.^[288-289] Eine Ausnahme bilden die sogenannten CpG-Inseln (CGIs). Dies sind DNA-Abschnitte mit einer durchschnittlichen Länge von 1000 Basenpaaren, welche eine deutlich erhöhte CpG-Dichte besitzen und die normalerweise hypomethyliert sind.^[19-20] Etwa 70% aller Genpromotoren sind gleichzeitig CGIs, d.h. ihre DNA-Sequenz enthält überdurchschnittlich viele CpG-Dinukleotide.^[21] Bei der Sequenzierung von hmC hat sich gezeigt, dass in Promotoren mit geringer bis mittlerer CpG-Dichte viele Cytosine hydroxymethyliert sind. Da die Expressionsrate eines Gens positiv mit der CpG-Dichte seines Promotors korreliert, sind diese Promotorregionen mit schwach bis mittelmäßig stark exprimierten Genen assoziiert.^[218-219, 221, 240] Die Promotorregionen stark exprimierter Gene enthalten dagegen kaum hmC und sind auch größtenteils unmethyliert.^[240, 283]

Vergleicht man die Ergebnisse der hmC-Sequenzierung von humanen und murinen Geweben, so lässt sich eine weitgehende Konservierung der hmC-Verteilung zwischen den Arten feststellen.^[278] Die dabei gefundene Verteilung von hmC im Genom der Säugetiere (siehe oben) lässt darauf schließen, dass diese Base in starkem Maße an der epigenetischen Kontrolle der Transkription beteiligt ist.^[49, 78] Die Akkumulation von hmC mit zunehmendem Alter^[115] wie auch Tracingexperimente mit stabil isotoopenmarkierten Nukleosiden^[40] deuten darauf hin, dass es sich bei einem Großteil der hmC-Basen im Genom um stabile Modifikationen handelt,^[290] welche die Genexpression unabhängig von ihrer Bedeutung als Intermediat der aktiven, d.h. enzymatischen DNA-Demethylierung beeinflussen können.

3.3.2 5-Formylcytosin (fC) und 5-Carboxycytosin (caC)

Formylcytosin und Carboxycytosin sind in deutlich geringeren Mengen im Genom der Säugetiere vorhanden als hmC. Die Base fC kommt hier etwa zehnmal häufiger vor als caC, trotzdem sind nur etwa 0,02-0,002% aller Cytosine formyliert.^[31] Ebenso wie bei hmC beobachtet man auch bei Formylcytosin eine ausgeprägte Abhängigkeit des globalen DNA-Modifikationsgrades vom Lebensalter. Während aber das hmC-Level bei Säugetieren mit zunehmendem Alter in Organen und Geweben ansteigt, nimmt das fC-Level schon direkt nach der Geburt und in den frühen

Entwicklungsstadien des Organismus stark ab (siehe Abbildung 12).^[115] Die höchsten endogenen Level an fC und caC findet man in embryonalen Stammzellen, während in adulten Geweben nur fC in geringen Mengen vorkommt und caC dort bisher nicht detektiert werden konnte.^[31] Eine genomweite Sequenzierung dieser Basen wurde deshalb zunächst in (murinen) embryonalen Stammzellen und neuralen Progenitorzellen (mESCs und mNPCs) vorgenommen.^[220, 229-230, 241, 263] Da fC und caC von der Glykosylase TDG aus der DNA herausgeschnitten werden können^[27, 34] (siehe Abschnitt 3.4.2.1), wurden neben dem Wildtyp auch Zellen mit Tdg-Knockout oder Knockdown sequenziert. Der dabei beobachtete Anstieg des fC-Levels um den Faktor 2-6 bzw. des caC-Levels um den Faktor 8 offenbart, dass diese Modifikationen in hohem Maße dynamisch reguliert sind.^[220, 229-230]

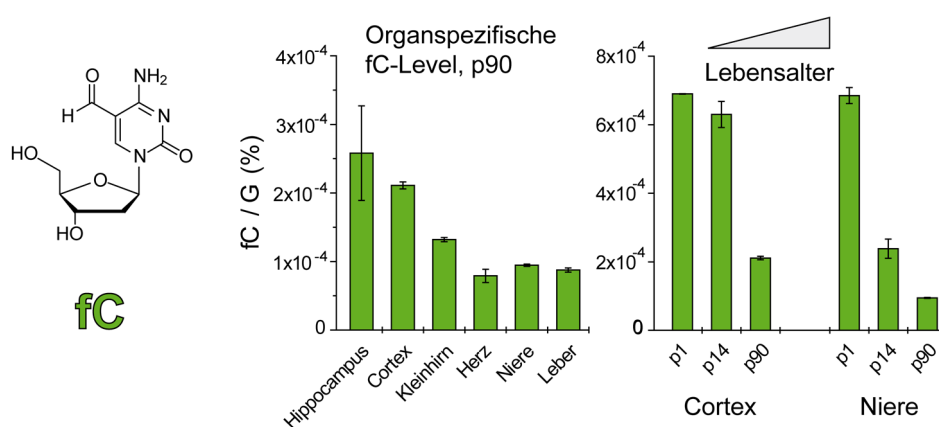


Abbildung 12. Organspezifische Level^[31] und altersabhängige Zunahme^[115] von fC in der Maus. Alle Werte sind als *Modifikationen pro hundert Guaninbasen* angegeben. Guanin (G) wurde als Referenz gewählt, weil die Zahl der Guaninbasen der Summe aus Cytosin (C) und den Cytosinderivaten mC, hmC und fC entspricht. p1 = postnataler Tag 1, d.h. erster Tag nach der Geburt, p14 und p90 entsprechend.

Betrachtet man die Verteilung von fC und caC im mESC-Genom, so zeigt sich, dass beide Basen ähnlich wie hmC hauptsächlich in Sequenzabschnitten mit regulatorischer Funktion und in den Genkörpern lokalisiert sind.^[49] fC weist dabei genau wie hmC eine starke Strangasymmetrie bezüglich der Formylierung der CpG-Sequenzpalindrome auf.^[241, 264] Besonders hoch ist die lokale Dichte von fC und caC in distalen regulatorischen Elementen. Hier findet man die Basen bevorzugt in aktiven und bivalenten Enhancern.^[220, 230, 291] Bivalente Enhancer sind zunächst inaktiv, sie können aber bei Bedarf sehr schnell in einen aktiven Status überführt werden. Möglicherweise also ermöglicht erst das Vorhandensein von fC und caC die rasche Aktivierbarkeit dieser Enhancer, da beide Basen jederzeit durch TDG entfernt und durch unmodifizierte Cytosine ersetzt werden können (für Details siehe Abschnitt 3.4.2.1).^[34] Ein Tdg-knockout in mESCs führt zu einer weiteren Akkumulation von fC und caC besonders in den aktiven Enhancern. Dies deutet darauf hin, dass permanente TET / TDG-vermittelte Demethylierungen in diesen Sequenzabschnitten notwendig sind, um eine

Hypermethylierung und damit eine Inaktivierung dieser Elemente zu verhindern (eine genaue Beschreibung der TET / TDG-vermittelten enzymatischen DNA-Demethylierung über die Intermediate fC und/oder caC findet sich in Abschnitt 3.4.2.1).^[220] Dasselbe gilt in ESCs offenbar auch für die Bindungsstellen pluripotenzassoziiierter Transkriptionsfaktoren. So führt ein Tdg-knockout in mESCs zu einer caC-Akkumulation an den Bindungsstellen von OCT4, NANOG, SOX2 und ESRRB, was auf eine verstärkte Demethylierungsaktivität in diesen Bereichen hindeutet.^[220]

Auch in proximalen Promotoren findet man die Basen fC und caC in erhöhter lokaler Konzentration.^[220, 229, 263] Eine Korrelation der lokalen fC-Dichte in diesen Promotoren mit der Transkriptionsaktivität der regulierten Gene liefert jedoch widersprüchliche Ergebnisse. Während Raiber *et al.*^[229] und Neri *et al.*^[263] besonders viel fC und caC in den hypomethylierten Promotoren stark exprimierter Gene fanden, wurden von Shen *et al.*^[220] erhöhte fC-Level in den Promotoren von Genen mit geringer bis mittlerer Aktivität gemessen.

Bei der Differenzierung von neuronalen Progenitorzellen (NPCs) zu Neuronen und Gliazellen kommt es zu einer transienten Akkumulation von fC und caC in zelltypspezifischen Promotoren.^[292] Diese Promotoren sind in neuronalen Progenitorzellen stark methyliert, so dass die entsprechenden zelltypspezifischen Gene inaktiv sind. Während der Differenzierung werden diese Gene durch Demethylierung der Promotoren aktiviert. Die Demethylierung erfolgt dabei offenbar über eine TET-vermittelte Oxidation von mC zu fC und caC, welche dann von TDG aus der DNA herausgeschnitten und anschließend durch Cytosin ersetzt werden können. Entsprechend führt ein TDG-*knockdown* während der Differenzierung zu einer zusätzlichen Akkumulation von fC und caC in diesen Promotoren.^[292]

In den Genkörpern findet man sowohl die Cytosinmodifikationen fC und caC, als auch mC und hmC in erhöhter Konzentration, jedoch ist die biologische Bedeutung der Basen in diesem Kontext bisher weitgehend ungeklärt.^[220, 229-230] Es wurde vorgeschlagen, dass die Cytosinmodifikationen im Genkörper an der Feinregulierung der Genaktivität beteiligt sind, etwa indem sie die Bindung zwischen der RNA-Polymerase II und der DNA beeinflussen.^[293-294] Möglicherweise sind die Modifikationen daneben an Splicingprozessen beteiligt, da die lokale Modifikationsdichte in Exons höher als in Introns ist.^[49] Natürlich ist nicht auszuschließen, dass fC und caC auch in den Genkörpern nur als Intermediate der aktiven Demethylierung *via* TET / TDG auftreten, um auch hier eine Hypermethylierung der DNA zu verhindern.

Wie oben beschrieben, haben die Cytosinmodifikationen fC und caC offenbar eine außerordentliche Bedeutung als Intermediate aktiver, d.h. enzymatischer DNA-Demethylierungsprozesse.^[295] Daneben besitzen die Basen fC und / oder caC möglicherweise auch eine Bedeutung als eigenständige

epigenetische DNA-Modifikationen. So wurden bei der Suche nach mit fC oder caC interagierenden Proteinen potentielle fC-spezifische Interaktoren entdeckt, die nicht zur Gruppe der DNA-Reparaturenzyme gehören (siehe auch Kapitel 4.10).^[41, 296] Auch haben Tracingexperimente mit stabil isotoopenmarkierten Nukleosiden gezeigt, dass zumindest ein Teil der fC-Basen nicht direkt durch TDG aus der DNA herausgeschnitten wird, sondern über längere Zeiträume in der DNA verbleibt.^[297] In diesem Zusammenhang ist es interessant, dass fC in bestimmten Sequenzabschnitten eine ähnliche lokale Konzentration wie mC oder hmC besitzt.^[241] Daneben wurde berichtet, dass fC-Basen die lokale Struktur der DNA verändern können.^[298] Dies hat möglicherweise einen direkten Einfluss auf die Transkription der entsprechenden DNA-Abschnitte.

3.3.3 5-Hydroxymethyluracil (hmU)

Wie Carell und Mitarbeiter gezeigt haben, kann neben hmC, fC und caC auch die Base hmU durch TET-vermittelte Oxidation gebildet werden (siehe auch Kapitel 4.9).^[31] Eine Quantifizierung der globalen hmU-Level in Mäusegeweben und murinen embryonalen Stammzellen (mESCs) ergab,^[31] dass das hmU-Level in mESCs mit $2,5 \times 10^{-4}\%$ hmU-Basen pro 100 Guanin-Basen in etwa so hoch wie das caC-Level ist (siehe Abbildung 13). Genau wie caC kommt auch hmU als enzymatisch gebildete DNA-Modifikation nur in embryonalen Stammzellen vor.^[31] Die (wenigen) in ausdifferenziertem Gewebe detektierten hmU-Basen sind im Wesentlichen das Resultat einer oxidativen DNA-Schädigung.^[117] Die biologische Bedeutung des enzymatisch generierten hmUs in embryonalen Stammzellen ist bisher unbekannt.^[299]

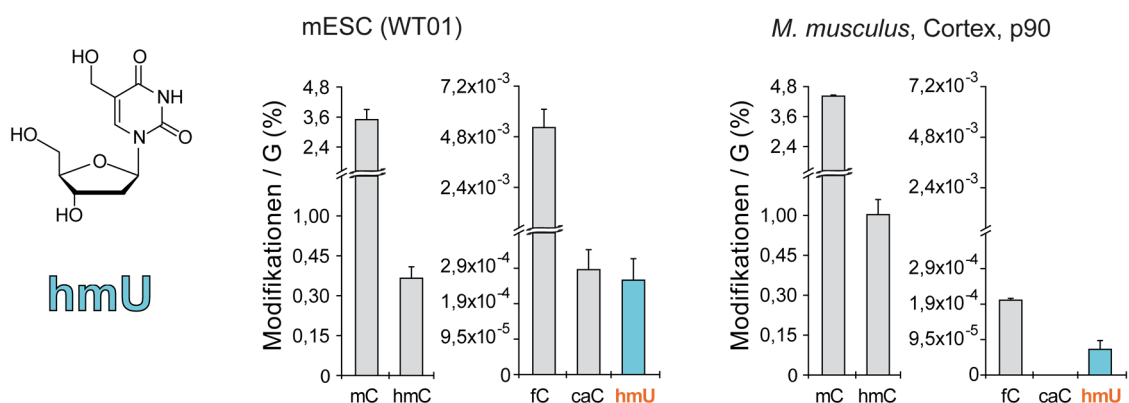


Abbildung 13. hmU-Level in embryonalen Stammzellen (mESC) des Typs WT01 und im Cortex (p 90, 90 Tage nach der Geburt) der Maus. Alle Werte sind als *Modifikationen pro hundert Guaninbasen* angegeben. Während hmU in den Stammzellen größtenteils enzymatisch gebildet wird, entsteht es im Cortex hauptsächlich durch oxidative DNA-Schädigung. Zum Vergleich sind neben hmU auch die mC-, hmC-, fC- und caC-Level abgebildet.

3.4 DNA-Demethylierung in Säugetieren

Zu bestimmten Zeitpunkten in der Entwicklung eines Organismus ist es notwendig, die DNA teilweise oder praktisch vollständig zu demethylieren.^[49] Dabei werden methylierte Cytosine im Genom durch unmethylierte Cytosine (C) ersetzt. Die im Methylierungsmuster enthaltene epigenetische Information wird dabei gelöscht und zuvor stummgeschaltete Gene können wieder exprimiert werden.^[36, 300-301] Wie in Kapitel 3.3 beschrieben, spielen die TET-Enzyme und die von ihnen generierten Basen hmC, fC und caC bei diesen Prozessen offenbar eine wesentliche Rolle.^[29, 47, 49, 78, 103, 278, 300-302]

Wie wir gleich sehen werden, treten bei Säugetieren sowohl globale^[106, 300, 303-304] als auch lokale^[158-163] DNA-Demethylierungsereignisse auf. Unter einer globalen DNA-Demethylierung versteht man dabei eine Demethylierung des Genoms, von der nur wenige DNA-Abschnitte ausgenommen sind. Dagegen bleibt bei lokalen DNA-Demethylierungen das Methylom größtenteils unverändert, es kommt nur an wenigen Positionen im Genom zu einem Austausch von mC durch C. Während globale DNA-Demethylierungen bei Säugetieren nur während der Embryogenese auftreten,^[106, 300, 303-304] besitzen lokale Demethylierungen bis ins hohe Lebensalter eine besondere Bedeutung für die neuronale Plastizität.^[158-163]

Mechanistisch lassen sich DNA-Demethylierungsereignisse grob in zwei Kategorien einteilen, und zwar in aktive und passive Demethylierungen. Unter einer aktiven DNA-Demethylierung versteht man allgemein eine enzymatische Transformation von mC zu Cytosin, welche nicht an die DNA-Replikation gekoppelt sein muss.^[36] Als passive Demethylierung bezeichnet man klassischerweise die Verdünnung von mC, die stattfindet, wenn bei der DNA-Replikation der neugebildete Gegenstrang nicht methyliert wird.^[305] Ein Spezialfall der passiven Demethylierung^[103] läge dann vor, wenn mC zunächst zu hmC, fC oder caC oxidiert werden würde und erst dann eine replikationsgekoppelte Verdünnung stattfindet.

Im Folgenden soll zunächst genauer erläutert werden, wann und wo globale und lokale DNA-Demethylierungen in Säugetieren auftreten. Mögliche Mechanismen der aktiven DNA-Demethylierung und die Beteiligung der TET-Enzyme bzw. der TET-generierten Cytosinmodifikationen hmC, fC und caC an diesen Prozessen werden dann in Abschnitt 3.4.2 ausführlich diskutiert.

3.4.1 Globale und lokale DNA-Demethylierungen in Säugetieren

3.4.1.1 Globale DNA-Demethylierungen während der Embryogenese

Während der Embryogenese von Säugetieren kommt es zweimal zu einer globalen Demethylierung der genomischen DNA, von der nur wenige DNA-Abschnitte ausgenommen sind^[106]. Die erste Demethylierungswelle tritt noch in der Zygote auf, es wird dabei die Totipotenz der Zelle etabliert.^[152] Später findet eine weitere globale Demethylierung in den sich entwickelnden Urkeimzellen (*primordial germ cells, PGCs*) statt.

Solange sich die Zellkerne von Eizelle und Spermium in der Zygote nicht vereinigt haben, liegen sie als räumlich getrennte Pronuklei vor. Noch vor der ersten Zellteilung beginnt im väterlichen Pronukleus die globale Demethylierung des Genoms.^[150] Dieser Prozess läuft so schnell ab, dass er nicht durch eine passive mC-Verdünnung bei der DNA-Replikation erklärt werden kann.^[306] Tatsächlich konnte durch Immunfärbung gezeigt werden, dass das mC im väterlichen Pronukleus zunächst enzymatisch zu hmC oxidiert wird.^[148] Möglicherweise findet teilweise auch eine Weiteroxidation zu fC und caC statt.^[149] Die Oxidation von mC erfolgt dabei durch TET3, welches im Gegensatz zu TET1 und TET2 in der Zygote stark exprimiert wird und im paternalen Pronukleus angereichert vorkommt.^[147] Entsprechend führt ein Tet3-*knockdown* oder *knockout* zu einer Störung der weiteren embryonalen Entwicklung, da die für die Totipotenz der Zygote entscheidenden Schlüsselgene Oct4 und Nanog im paternalen Pronukleus nicht rechtzeitig demethyliert werden können.^[147, 307] Während der auf die mC-Oxidation folgenden Zellteilungen kommt es dann zu einer passiven Verdünnung von hmC (und evtl. fC und caC),^[308] da hemihydroxymethylierte Positionen in der DNA im Gegensatz zu hemimethylierten Positionen von der *Maintenance*-DNA-Methyltransferase DNMT1 nicht als Templat erkannt werden.^[309] Nach der von Zhang vorgeschlagenen Terminologie handelt es sich daher insgesamt um eine DNA-Demethylierung durch aktive Modifizierung (von mC) und passive Verdünnung (von hmC).^[103]

Auch das mütterliche Genom wird während der frühen Embryogenese global demethyliert. Die Demethylierung erfolgt hier durch passive Verdünnung von mC, d.h. bei der DNA-Replikation erfolgt keine Übertragung des Methylierungsmusters auf den neugebildeten Gegenstrang.^[310] Es ist interessant, dass trotz starker TET3-Expression in der Zygote im mütterlichen Genom keine Oxidation von mC zu hmC (und/oder fC und caC) stattfindet. Offenbar wird hier die DNA aktiv vor einer mC-Oxidation geschützt. Verantwortlich dafür ist das Enzym STELLA (auch bekannt als DPPA3 und PGC7), welches im Zusammenspiel mit der Histonmodifikation H3K9me2 einen TET3-Zugriff auf die

DNA verhindert.^[311-312] Es wurde beobachtet, dass ein *Stella-knockout* hinreichend ist, um auch im mütterlichen Pronukleus eine Oxidation von mC zu hmC auszulösen.^[307]

Welchen biologischen Vorteil die beobachtete Asymmetrie bei der Demethylierung des mütterlichen und väterlichen Genoms hat, ist bisher nicht bekannt. Möglicherweise ist es in diesem Zusammenhang von Bedeutung, dass die DNA eines Spermiums im Vergleich zur DNA einer Eizelle hypermethyliert ist (90 % aller CpG-Dinukleotide liegen in Spermien methyliert vor, während in Eizellen nur 40 % aller CpGs methyliert sind).^[313] Dies erfordert eventuell eine rasche Angleichung der Methylierungslevel durch mC-Oxidation im väterlichen Pronukleus der Zygote.

Eine globale Demethylierung des Genoms findet in Säugetieren auch bei der Bildung der Urkeimzellen (*primordial germ cells*, PGC) aus den Epiblastzellen statt, um so deren Pluripotenz zu etablieren.^[106, 154-155, 157, 304, 314] Diese Demethylierung erfolgt in zwei aufeinanderfolgenden Wellen. Während die sich entwickelnden PGCs in Richtung der Gonaden wandern, kommt es zunächst zu einer passiven mC-Verdünnung.^[315-316] Anschließend kommt es zu einem raschen mC-Verlust durch aktive Prozesse.^[155, 157, 304, 317]

Die erste Demethylierungswelle wird eingeleitet durch eine Ummodellierung der Chromatinstruktur. Charakteristisch dafür ist der Rückgang der Histonmodifikation H3K9me2 bei gleichzeitigem Anstieg von H3K27me3.^[318-319] H3K27me3 ist genau wie H3K9me2 eine Modifikation, welche die Transkription unterdrückt. Offenbar ermöglicht sie aber im Unterschied zu H3K9me2 – wenn die DNA-Demethylierung genügend fortgeschritten ist – einen schnelleren Neustart der Transkription.^[320-321] Die Reorganisation des Chromatins wird begleitet von einer verminderten Expression der *de-novo*-Methyltransferasen und von UHRF1, welches zwingend notwendig für die Rekrutierung der *Maintenance*-DNA-Methyltransferase DNMT1 ist.^[322-323] Das Resultat ist eine breite Demethylierung des Genoms durch passive mC-Verdünnung bei der Zellteilung.^[315-316, 318]

Nachdem sich die PGCs in den Gonaden eingenistet haben, findet eine rasche Demethylierung weiterer DNA-Abschnitte statt.^[155, 157, 304, 317, 319] Die Kinetik des mC-Verlustes deutet darauf hin, dass es sich hierbei um einen aktiven Prozess handeln muss,^[154, 317] die Demethylierung erfolgt dabei wahrscheinlich über eine TET1-vermittelte Oxidation von mC.^[324] Ein *Tet1-Knockout* in PGCs der Maus führt deshalb zu einer Störung der Meiose und einer starken Verminderung der Fertilität.

Das hier beschriebene Zusammenspiel von passiver und aktiver Demethylierung bei der Urkeimzellbildung ist bisher noch nicht vollständig verstanden. Es ist aber offenbar die Voraussetzung für die akkurate und zeitlich genau koordinierte Etablierung der Pluripotenz der Gameten.

3.4.1.2 Lokusspezifische, aktive DNA-Demethylierung in somatischen Zellen

Erst seit wenigen Jahren gibt es robuste experimentelle Hinweise darauf, dass in somatischen Zellen aktive DNA-Demethylierungen auftreten können.^[325-331] Diese erfolgen lokusspezifisch, z.B. nur in der Promotorregion bestimmter Gene, und der genaue Reaktionsweg der Demethylierung ist in manchen Fällen nicht abschließend geklärt.^[36, 49, 78, 300-302]

Postmitotische Neuronen sind *in vitro* und *in vivo* ein nahezu ideales Modellsystem für das Studium der aktiven DNA-Demethylierung. Sie teilen sich nicht mehr, daher muss eine eventuell auftretende Demethylierung ihrer DNA auf aktivem Wege erfolgen. In den letzten Jahren hat sich gezeigt, dass in Neuronen tatsächlich lokusspezifische Demethylierungen auftreten.^[158, 332-334] Sie sind die direkte Folge externer Stimuli, welche die neuronale Aktivität anregen. Demethylierungen treten dabei in der Promotorregion von Genen wie *Bdnf* auf, die mit neuronaler Plastizität und dem Lernvermögen assoziiert werden.^[158, 333] *Bdnf* (*Brain-derived neurotrophic factor*) ist als Neurotrophin von außerordentlicher Wichtigkeit für ein funktionierendes Langzeitgedächtnis und die neuronale Plastizität des adulten Hirns.^[333, 335] Besonders gut untersucht ist die aktive Demethylierung der *Bdnf*-Promotorregion IX.^[327] Sie erfolgt u.a. als Reaktion auf eine elektrokonvulsive Stimulation (ECS) der neuronalen Aktivität im Hippocampus der Maus.^[33, 336] Die BDNF-Expression als Folge der Promotordemethylierung fördert dabei die Neurogenese im adulten Hirn. Die Demethylierung des *Bdnf*-Promotors IX im Hippocampus der Maus beginnt mit einer TET1-katalysierten Oxidation von mC zu hmC (und evtl. auch zu fC und / oder caC).^[33] Wird TET1 im Hippocampus durch einen shRNA-vermittelten *in-vivo-knockdown* gezielt herunterreguliert, findet auch bei einer Stimulation der neuronalen Aktivität keine Demethylierung dieses Promotors mehr statt.^[33]

Aktive DNA-Demethylierungen konnten auch als Folge von weitaus komplexeren Stimuli nachgewiesen werden, u.a. im Hippocampus.^[158] Dieser ist von grundlegender Bedeutung für die Überführung von Gedächtnisinhalten aus dem Kurzzeit- in das Langzeitgedächtnis. Wird beispielsweise bei Mäusen eine kontextspezifische Angstkonditionierung (*contextual fear conditioning*) durchgeführt, so kommt es im Hippocampus zu lokusspezifischen Demethylierungen in der Promotorregion von *Bdnf*.^[329] Auch in der Promotorregion von *reelin*, einem Schlüsselgen der synaptischen Plastizität, konnten solche Demethylierungen nachgewiesen werden.^[337] Eine erhöhte Expression dieser Gene als Resultat der Demethylierungen ist direkt mit der Fähigkeit verknüpft, die bei der Angstkonditionierung erlernten Erinnerungen später wieder abzurufen.

Tatsächlich ist die epigenetische Basis für das Speichern von neuen Gedächtnisinhalten ein komplexes Zusammenspiel aus DNA-Demethylierungs- und *de-novo*-Methylierungsereignissen.^[158] So findet neben den Demethylierungsereignissen eine Promotormethylierung und damit eine verminderte

Expression von Genen statt, die dem Erlernen von neuen Inhalten entgegenarbeiten. Dies ist z.B. beim protein phosphatase 1-Gen (pp1) der Fall, welches der Erinnerungskonsolidierung entgegenwirkt.^[337] Grundsätzlich scheint das neuronale DNA-Methylom weitaus stärker und flexibler auf eine Stimulation der neuronalen Aktivität zu reagieren, als noch vor kurzem vermutet wurde. So wurde von Guo *et al.* durch *Next-generation Sequencing* gezeigt, dass sich in den Neuronen des Mäusehippocampus der Methylierungsstatus bei ca. 1,4 % aller CpGs ändert, wenn eine synchrone Stimulation der Neuronen durchgeführt wird.^[338] Es kommt dabei sowohl zu aktiven CpG-Demethylierungen als auch zu *de-novo*-Methylierungen. Sie treten hauptsächlich in Genen auf, welche die neuronale Plastizität beeinflussen. Die als Folge von neuronaler Aktivität auftretenden Änderungen des DNA-Methyloms werden generell von ebenso ausgeprägten Änderungen auf der Ebene der Histonmodifikationen begleitet.^[339]

Außer in Neuronen sind aktive DNA-Demethylierungen auch in anderen somatischen Zelltypen nachgewiesen worden. So zeigen humane Brustkrebszellen eine östrogenabhängige periodische De- und Remethylierung des pS2/TFF1-Promotors, welche mit einer periodischen Schwankung des entsprechenden mRNA-Levels einhergeht.^[330] Dieser Prozess wird initiiert, indem das Östrogen Estradiol den Transkriptionsfaktor Estrogen Factor α (EF α) aktiviert. EF α bindet an den pS2-Promotor und veranlasst ein periodisches Durchlaufen des Transkriptionszyklus. Dieser beginnt jeweils mit der Demethylierung des Promotors und endet mit dessen Remethylierung. Die Demethylierung muss dabei nach einem aktiven Mechanismus erfolgen, weil sie unabhängig von der Zellteilung stattfindet. Eine vergleichbare zyklische De- und Remethylierung des Promotors wurde auch bei anderen Genen gefunden, bei denen eine periodische Schwankung des transkribierten mRNA-Levels beobachtet wird.^[328] Dies deutet darauf hin, dass die dynamische Regulation der Genexpression mit einer entsprechenden dynamischen Veränderung des Methylierungsstatus der DNA einhergehen kann. Damit verbundene aktive und lokusspezifische DNA-Demethylierungen finden also möglicherweise häufiger statt als bisher vermutet.

Was ist nun der genaue Mechanismus der hier beschriebenen aktiven DNA-Demethylierungen? Um diese Frage –soweit zurzeit möglich– zu beantworten, werden im folgenden Kapitel mögliche Mechanismen der aktiven DNA-Demethylierung vorgestellt und ihre biologische Relevanz anhand des aktuellen Kenntnisstandes diskutiert.

3.4.2 Mögliche Mechanismen der aktiven DNA-Demethylierung

Die DNA-Methylierung als Mittel zur langfristigen epigenetischen Fixierung des Transkriptoms und damit natürlich auch des Proteoms einer Zelle^[340] ist bereits seit Jahrzehnten bekannt.^[341] Auch die während der Embryogenese auftretenden globalen DNA-Demethylierungswellen^[106, 157] kennt man seit langem.^[342] Diese Demethylierungswellen können jedoch auch mit einer passiven mC-Verdünnung bei der DNA-Replikation und Zellteilung erklärt werden.^[305, 316] Erst die genaue Untersuchung postmitotischer Neuronen, bei denen keine Zellteilung mehr stattfindet, lieferte 2003 den endgültigen Beweis für die Existenz von aktiven, d.h. enzymatischen DNA-Demethylierungsprozessen in Säugetieren.^[326-327] Zunächst war es allerdings nicht klar, nach welchem Mechanismus diese Demethylierungen ablaufen und welche Enzyme daran beteiligt sind.^[36, 343] Erst die Aufklärung der Funktion der TET-Enzyme^[23, 101] und die „Wiederentdeckung“ von hmC^[22-24] brachten neue Bewegung in dieses Forschungsfeld. Schnell war klar, dass die TET-Enzyme eine Schlüsselrolle bei der aktiven DNA-Demethylierung einnehmen,^[33, 109] und schon von Tahiliani *et al.*^[23] und Kriaucionis *et al.*^[22] wurde spekuliert, dass es sich bei hmC um ein Intermediat dieser Demethylierung handeln könnte. Im Rahmen der intensiven Forschung auf diesem Gebiet stellte sich jedoch heraus, dass der Prozess der aktiven DNA-Demethylierung komplizierter ist, als zunächst vermutet worden war. So besitzt hmC auch eine Rolle als eigenständige epigenetische DNA-Modifikation, welche unabhängig von der DNA-Demethylierung ist.^[40-41, 296] Außerdem konnte gezeigt werden, dass die TET-Enzyme nicht nur für die Bildung von hmC aus mC^[23] verantwortlich sind, sondern darüber hinaus auch die DNA-Basen fC,^[25-26] caC^[26-27] und hmU^[31] generieren können.

Welche Rolle diese neuentdeckten Basen im Prozess der aktiven DNA-Demethylierung einnehmen, welche Mechanismen für die aktive Demethylierung vorgeschlagen wurden und was der derzeitige Konsens bezüglich des tatsächlichen Mechanismus der aktiven DNA-Demethylierung ist, soll in diesem Abschnitt diskutiert werden.^[29, 36, 47, 49, 78, 103, 278, 300-302, 344-346]

Betrachtet man die Gesamtheit der bis jetzt bekannten, enzymatisch generierten modifizierten DNA-Basen, so lassen sich mehrere aktive DNA-Demethylierungsszenarien postulieren, bei denen diese Basen eine Rolle spielen. Die vorgeschlagenen Mechanismen lassen sich dabei grundsätzlich in zwei Gruppen einteilen, die sich darin unterscheiden, ob das gesamte Nukleotid oder nur die C5-Position der Base angegriffen wird.^[36, 49] Bei Letzteren wird im Schlüsselschritt die Bindung zwischen der Cytosinbase und der (substituierten) Methylgruppe an Position 5 gespalten. Auf diese Weise entsteht aus der modifizierten Base wieder das ursprüngliche Cytosin, ohne dass es zu einem Austausch der gesamten Base kommt. Bei den anderen vorgeschlagenen Mechanismen besteht der entscheidende Schritt jeweils aus der Entfernung der gesamten Base durch eine Glycosylase. Dabei wird die *N*-

glycosidische Bindung zwischen der modifizierten Base und dem Zucker gespalten. Das Resultat ist eine abasische Stelle in der DNA, die dann im Rahmen der Basenexzisionsreparatur (BER) prozessiert wird.^[347-349]

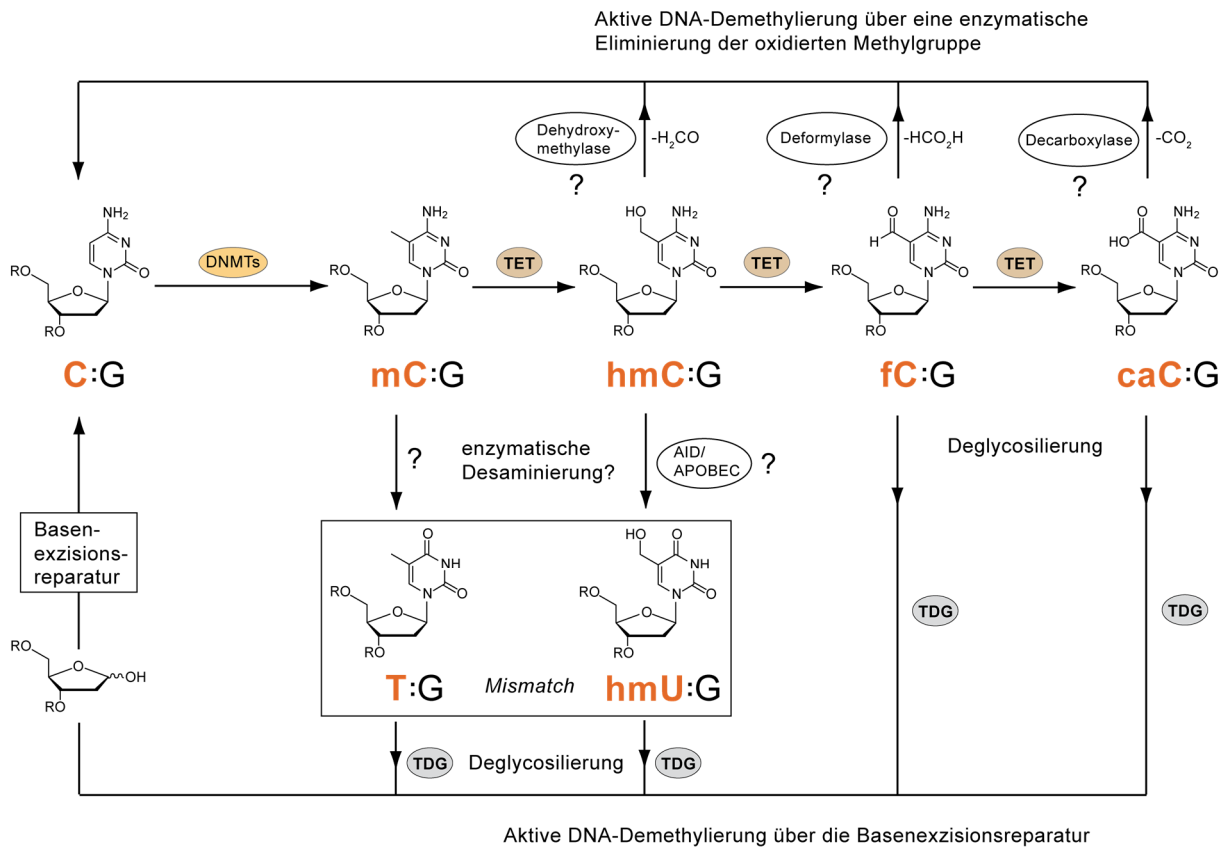


Abbildung 14. Mögliche Wege der aktiven DNA-Demethylierung, welche nach der Aufklärung der enzymatischen Aktivität der TET-Enzyme postuliert worden sind.

Bei der Basenexzisionsreparatur handelt es sich um einen Mechanismus zur Reparatur beschädigter (d.h. von oxidierten, desaminierten etc.) DNA-Basen.^[347-349] Dabei wird die beschädigte Base zunächst durch eine DNA-Glykosylase aus der DNA herausgeschnitten. Eine Endonuklease entfernt anschließend den verbliebenen Zuckerrest und evtl. umliegende Nukleoside. Schließlich werden die fehlenden Nukleoside durch eine DNA-Polymerase mit dem ungeschädigten Gegenstrang als Templat ersetzt. Eine aktive Demethylierung könnte daher erfolgen, indem mC (oder ein oxidiertes mC-Derivat) ähnlich wie ein DNA-Schaden prozessiert wird.

Methylcytosin oder die TET-generierten Basen hmC, fC und caC sind jedoch nicht die einzigen möglichen Substrate einer BER-basierten aktiven DNA-Demethylierung. Es wurden auch mehrstufige Mechanismen vorgeschlagen, bei denen zunächst eine Desaminierung der modifizierten Cytosinbase stattfindet.^[49, 302] Dabei würde mC in Thymin umgewandelt werden; aus hmC (bzw. fC oder caC) würde das entsprechenden Uracilderivat Hydroxymethyluracil (hmU) (bzw. Formyluracil (fU) oder

Carboxyuracil (caU)) gebildet werden, welches auch als BER-Substrat in Frage kommen könnte. Eine Übersicht über mögliche Wege der aktiven DNA-Demethylierung findet sich in Abbildung 14.

3.4.2.1 Aktive DNA-Demethylierung über die Glykosylase TDG im Rahmen der Basenexzisionsreparatur

Säugetiere besitzen (mindestens) elf verschiedene DNA-Glykosylasen.^[348] Von diesen besitzt die Glykosylase TDG für die aktive DNA-Demethylierung unter Beteiligung der Basenexzisionsreparatur (BER) nach derzeitigem Konsens eine besondere Bedeutung.^[49, 350] TDG (Thymin DNA-Glykosylase) gehört zur Familie der Uracil-DNA-Glykosylasen (UDGs),^[351] einer Gruppe von DNA-Reparaturenzymen, und kann Thymin und einige T-Derivate aus der DNA herausschneiden, wenn diese mit Guanin fehlgepaart sind.^[352] Die Hauptfunktion des Proteins wurde daher lange in der Verhinderung der Mutagenität der Desaminierung von mC zu T gesehen.^[353] Im Gegensatz zu den anderen Enzymen der UDG-Familie interagiert TDG jedoch auch mit Transkriptionsfaktoren,^[354-355] den *de-novo*-DNA-Methyltransferasen^[356] und histonmodifizierenden Enzymen.^[357] Dies legte eine weitergehende Beteiligung von TDG an epigenetischen Prozessen nahe und führte zu einer eingehenderen Untersuchung seiner biologischen Bedeutung.

Es zeigte sich, dass Tdg-defiziente Mäuse nicht lebensfähig sind und der Tdg-Doppelknockout (Doppelknockout = Knockout beider Allele) schon am 10.Tag nach der Befruchtung zum Absterben der Mäuseembryonen führt.^[35, 358] Zelllinien, die aus diesen Embryonen gewonnen wurden, weisen eine ausgeprägte Dysregulation der Genexpression auf. Die betroffenen Gene sind dabei in ihrer Promotorregion hypermethyliert. Das Gleiche gilt auch für murine embryonale Stammzellen (mESCs) mit Tdg-Doppelknockout, die mit Hilfe von Retinsäure zur Differenzierung angeregt werden.^[229] Die Hypermethylierung führt hier zu einem Absterben der Zellen, bevor diese vollständig differenziert sind. Die im Zusammenhang mit dem TDG-Knockout beobachtete Hypermethylierung tritt auch dann auf, wenn TDG als katalytisch inaktive Mutante vorliegt.^[35] Dies bedeutet, dass TDG mit seiner Glykosylaseaktivität direkt für ein korrektes Methylierungsmuster des Genoms sorgt.

Was ist nun das Substrat von TDG? Man könnte zunächst meinen, dass es sich um mC handeln muss,^[359] da in Pflanzen mC-Glykosylasen existieren, welche dort die aktive DNA-Demethylierung initiieren.^[360] Tatsächlich jedoch weisen weder TDG noch eine der anderen bekannten DNA-Glykosylasen von Säugetieren eine Glykosylaseaktivität bezüglich mC auf.^[352] Erst die Entdeckung der höher oxidierten mC-Derivate lieferte die Antwort auf diese Frage. Während neben mC auch hmC vom Enzym nicht als Substrat erkannt wird, werden sowohl fC als auch caC von TDG aus der DNA

herausgeschnitten.^[27, 34, 295, 350, 361-362] Die Aktivität von TDG bezüglich fC ist dabei, verglichen mit caC, etwa 4-5-mal höher.^[34]

Insgesamt ergibt sich folgender, experimentell zurzeit am besten belegter und weithin akzeptierter Mechanismus für die aktive DNA-Demethylierung in Säugetieren (siehe Abbildung 15):^[49] Methylierte Cytosine werden von den TET-Enzymen^[23, 109] über hmC^[22, 24] zunächst zu fC^[25, 27] und / oder caC^[26-27] oxidiert. fC und caC werden anschließend von der Glykosylase TDG aus der DNA herausgeschnitten,^[27, 34] wobei abasische Stellen in der DNA zurückbleiben. Diese werden dann im Rahmen der Basenexzisionsreparatur (BER) prozessiert und dabei durch unmodifizierte Cytosine ersetzt.^[348]

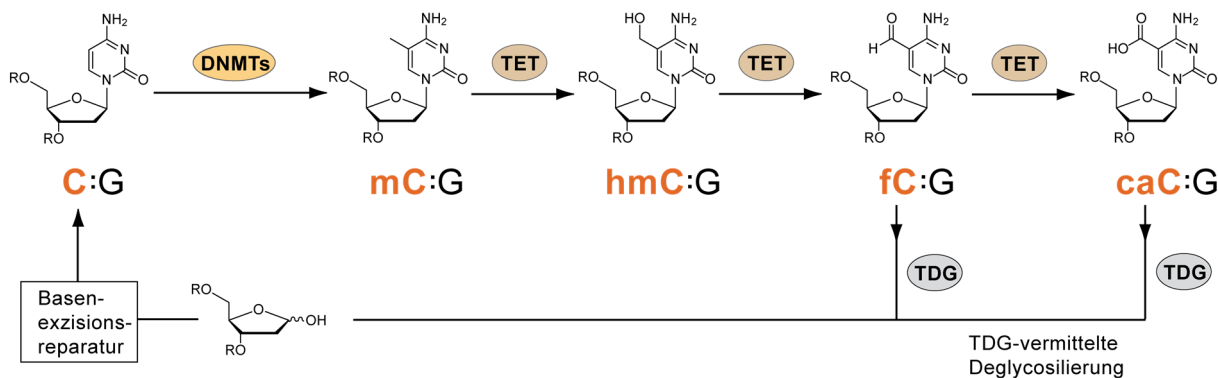


Abbildung 15. Aktive DNA-Demethylierung durch TET-katalysierte Oxidation von mC zu fC bzw. caC und TDG-vermittelte Basenexzisionsreparatur.

3.4.2.2 Demethylierung via Desaminierung

Der oben beschriebene Mechanismus der aktiven DNA-Demethylierung ist bisher experimentell am besten belegt. Es gibt jedoch Hinweise darauf, dass in manchen Fällen auch andere Routen der aktiven DNA-Demethylierung beschritten werden können.^[49]

So wurden mehrstufige, basenexzisionsreparatur-gekoppelte DNA-Demethylierungsmechanismen vorgeschlagen, welche als Schlüsselschritt eine enzymatische Desaminierung enthalten.^[49, 103] Das Substrat dieser Desaminierung könnte im einfachsten Falle Methylcytosin sein,^[363] aber auch hmC kommt als mögliches Substrat infrage. Durch Desaminierung wird mC in Thymin (T) umgewandelt; aus hmC entsteht das entsprechende Uracilderivat 5-Hydroxymethyluracil (hmU). In der DNA würde dies zu einer Fehlpaarung zwischen der desaminierten Base und dem gegenüberliegenden Guanin (G) führen (T:G bzw. hmU:G). Auf diese Weise könnte die entsprechende Base für die Entfernung aus der DNA markiert werden. Die Base könnte dann durch eine Glykosylase aus der DNA

herausgeschnitten und im Rahmen einer Basenexzisionsreparatur durch ein unmodifiziertes Cytosin ersetzt werden (siehe Abbildung 16).

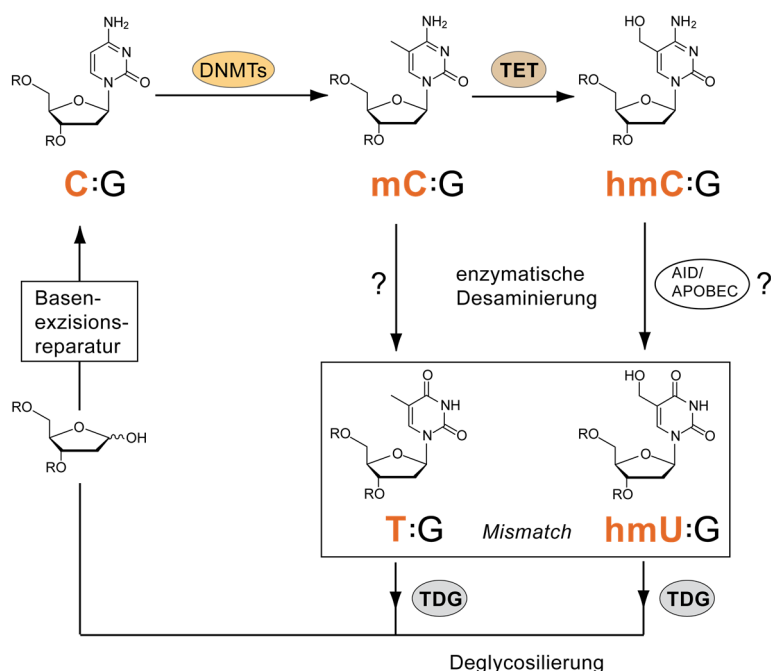


Abbildung 16. Postulierter Mechanismus der aktiven DNA-Demethylierung durch Desaminierung von mC bzw. hmC und TDG-vermittelte Basenexzisionsreparatur.

Auf der Suche nach einer Desaminase,^[364] welche die Reaktion von mC zu T katalysieren kann, wurden die Mitglieder der AID/APOBEC-Familie (AID = *activation-induced cytidine deaminase*; APOBEC = *apolipoprotein B mRNA-editing enzyme catalytic polypeptide*) besonders intensiv untersucht.^[363, 365-368] Diese Enzymfamilie katalysiert die Desaminierung von Cytosin zu Uracil in einzelsträngiger DNA und in der mRNA. APOBEC-3 beispielsweise kann Retrotransposons desaminieren und damit die Vermehrung von Retroviren hemmen.^[367, 369] APOBEC-1 ist an der Editierung von mRNA beteiligt und AID ist entscheidend an der Reifung von hochspezifischen Antikörpern beteiligt.^[370-372]

Ein Artikel von Rai *et al.* legt nahe, dass AID außerdem diejenige Desaminase sein könnte, welche in Zebrafischembryos durch Desaminierung von mC die aktive Demethylierung des Genoms einleitet.^[373] Laut dieser Studie ist die Glykosylase MBD4 verantwortlich für das darauffolgende Herausschneiden des fehlgepaarten T aus der DNA. Die Interaktion von AID und MBD4 wird dabei über das Protein GADD45 vermittelt. Auch in den Urkeimzellen der Maus (englisch *primordial germ cells*, PGCs) wurde eine Verbindung zwischen AID und der DNA-Demethylierung postuliert.^[374] In diesen Zellen führt ein Doppelknockout von AID zu einer Störung des genomweiten Methylierungsmusters, sie sind im Vergleich zum Wildtyp stark hypermethyliert. Allerdings gibt es auch ernsthafte Hinweise, die

gegen eine Desaminierung von mC durch AID/APOBEC sprechen.^[343] So sind Mäuse mit AID-Doppelknockout lebensfähig.^[375] Außerdem haben kinetische Studien ergeben, dass die Desaminaseaktivität der AID/APOBEC-Enzyme wesentlich vom sterischen Anspruch des Substituenten an Position 5 der Base abhängt.^[376-377] Je größer dieser Substituent ist, umso langsamer erfolgt die Desaminierung. Daher wird mC im Vergleich zu C, dem kanonischen Substrat, etwa zehnmal langsamer desaminiert.^[376-377] Berücksichtigt man zusätzlich, dass das Verhältnis von mC zu C in der DNA von Säugetieren etwa 1:20 beträgt, erscheint die Beteiligung der AID/APOBEC-Enzymfamilie an einer Desaminierung von mC eher unwahrscheinlich.^[18]

Neben den Enzymen der AID/APOBEC-Familie gibt es auch andere Kandidaten für die postulierte mC-Desaminase. Besonders interessant sind in diesem Zusammenhang die DNA *de-novo*-Methyltransferasen (DNMT3a/b). Es konnte gezeigt werden, dass DNMT3a und b *in vitro* und in Abwesenheit ihres Cofaktors S-Adenosylmethionin (SAM) mC zu T desaminieren können.^[330] Ob Dnmt3a oder Dnmt3b auch *in vivo* als mC-Desaminasen fungieren, ist derzeit noch unklar, besonders da ihr Cofaktor SAM in relativ hoher Konzentration in Zellen vorliegt und auch als Methylgruppendonor essentiell für das Funktionieren weiterer Enzyme ist.^[378-379]

Ein weiterer möglicher Weg der aktiven DNA-Demethylierung verläuft über die Desaminierung von hmC unter Bildung von Hydroxymethyluracil. Auch hier wurden die Enzyme der AID/APOBEC-Familie als potentielle Desaminasen genannt. Beispielsweise führt laut Guo *et al.* eine Überexpression von AID oder anderen Mitgliedern der AID/APOBEC-Familie zu einer Abnahme des durch TET1-Koexpression erzeugten hmC-Levels in HEK-Zellen.^[33] Auch im Hippocampus von Mäusen kommt es bei Überexpression von AID zu einer Reduktion des endogenen hmC-Levels.^[33] Cortellino *et al.* berichten, dass sich in Säugetierzellen ein ternärer Komplex aus AID, der Glykosylase TDG und dem Protein GADD45 bildet.^[35] TDG erkennt dabei hmU als Substrat, wenn es sich in einem hmU:G-Mismatch-Basenpaar befindet, und kann dieses aus der DNA entfernen. GADD45 vermittelt, wie auch in Zebrafischembryos gefunden, die Bindung zwischen AID und TDG und ist eventuell auch für die korrekte Bindung zur DNA verantwortlich.^[373]

Trotz dieser Ergebnisse erscheint die Desaminierung von hmC zu hmU als Hauptweg der aktiven DNA-Methylierung wenig wahrscheinlich. Dagegen sprechen zunächst einmal die schon oben erwähnten Arbeiten, nach denen die Desaminaseaktivität der AID/APOBEC-Enzyme mit zunehmender Größe des Substituenten an Position 5 abnimmt.^[376-377] Die Desaminaseaktivität von AID ist daher gegenüber hmC sogar noch geringer als gegenüber mC. Weiterhin zeigen auch die Ergebnisse von Carell und Mitarbeitern,^[31] dass das *Steady-State*-Level von hmU zumindest in mESCs nicht durch hmC-Desaminierung entsteht. Durch Isotopenmarkierungsexperimente konnten Carell und Mitarbeiter nachweisen, dass hmU stattdessen durch Oxidation von Thymin gebildet wird (siehe auch

chemisch sinnvollste Intermediat der aktiven DNA-Demethylierung durch Eliminierung des Substituenten. Bei den von Carell und Mitarbeitern durchgeführten Experimenten soll das Thiol eine Aminosäure der hypothetischen Decarboxylase imitieren, welche die Spaltung der C-C-Bindung katalysiert. Der postulierte Mechanismus dieser enzymatischen caC-Decarboxylierung ist in Abbildung 18 dargestellt.^[380]

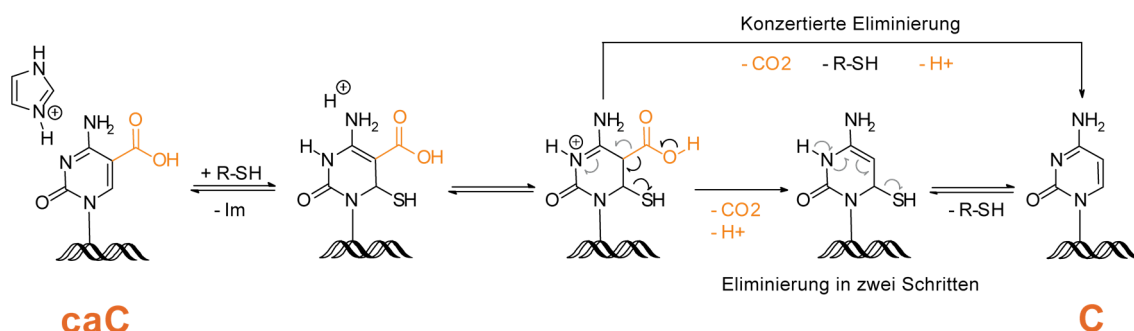


Abbildung 18. Von Carell und Mitarbeitern postulierter Mechanismus der enzymatischen caC-Decarboxylierung.^[380] Das Thiol initiiert bei der untersuchten Decarboxylierung eine Aminosäure der hypothetischen Decarboxylase und ist von entscheidender Bedeutung für die Reaktion. Es greift das caC in einer nukleophilen Addition an Position 6 an und sättigt so die C5-C6-Bindung des Aromaten, was durch die gleichzeitige Protonierung des Stickstoffs an Position 3 erleichtert wird. Im nächsten Reaktionsschritt erfolgt die Eliminierung von CO₂, gleichzeitig (oder auch im nächsten Schritt; graue Pfeile) wird das Thiol unter Rückbildung des aromatischen Systems abgespalten.

Die Experimente von Carell und Mitarbeitern zeigen, dass ein C-C-Bindungsbruch *in vitro* an der gewünschten Position prinzipiell möglich ist.^[380] Tatsächlich kennt man bereits eine ganz ähnliche Reaktion, die Teil des zellulären Recyclingwegs von Thymin ist. Bei dieser Reaktion wird zunächst die Methylgruppe einer freien Thymin-Base durch das Enzym Thymin-7-hydroxylase oxidiert.^[381] Vergleichbar mit der Oxidation von Methylcytosin über hmC und fC zu caC, erfolgt auch die Thymin-oxidation iterativ. Dabei entsteht über 5-Hydroxymethyluracil (hmU) und 5-Formyluracil (fU) 5-Carboxyuracil (caU). caU wird dann durch das Enzym Isoorotat-Decarboxylase (IDCase) decarboxyliert.^[382] Um eine entsprechende enzymatische Decarboxylierung für die DNA-Base Carboxycytosin nachzuweisen, wurde von Carell und Mitarbeitern schweratommarkiertes caC in Oligonukleotide inkorporiert.^[383] Diese wurden dann mit einem Kernextrakt aus murinen embryonalen Stammzellen inkubiert. Nach Re-isolation der Oligonukleotide wurde die Probe massenspektrometrisch untersucht. Dabei konnte auch eine geringe Menge schweratommarkiertes Cytosin detektiert werden, welches nur durch Decarboxylierung des schweratommarkierten caC entstanden sein kann.^[383] Ob an dieser Decarboxylierung allerdings eine caC-spezifische Decarboxylase beteiligt war, ließ sich durch dieses Experiment nicht ermitteln. Auch war die Menge des detektierten schweratommarkierten Cytosins so gering, dass auch eine unspezifische

Decarboxylierung möglich erscheint. Eine caC-spezifische Decarboxylase konnte bisher trotz intensiver Suche noch immer nicht identifiziert werden.

3.5 hmC-, fC- und caC-bindende Proteine als Übersetzer der epigenetischen Information

Die Bedeutung der oxidierten Methylcytosine für die DNA-Demethylierung ist durch eine Vielzahl von Experimenten sicher belegt.^[49] Daneben deutet vieles darauf hin, dass neben mC auch hmC eine Cytosinmodifikation mit eigenständiger epigenetischer Bedeutung ist.^[40-41, 296] Allerdings beginnt man erst langsam zu verstehen, wie die durch die Verteilung der hmC-Basen kodierte Information in eine biologische Wirkung übersetzt wird. So sind prinzipiell zwei Möglichkeiten denkbar, wie die Bindungsaffinität von DNA-bindenden Proteinen durch hmC beeinflusst werden kann. Zum einen könnte eine Hydroxymethylierung die Bindung bestimmter Proteine verhindern, welche eine Affinität für methylierte DNA-Abschnitte besitzen, und damit die Transkription beeinflussen. Zum anderen könnten hmC-spezifische Reader existieren, welche gezielt an hmC-haltige Sequenzabschnitte binden. Dabei könnte es sich z.B. um Transkriptionsfaktoren oder histonmodifizierende Enzyme handeln. Im folgenden Abschnitt sollen der derzeitige Kenntnisstand und die verfügbaren experimentellen Befunde kurz dargestellt werden.^[47, 106, 159-160, 302, 384-386]

3.5.1 Einfluss von Cytosinmodifikationen in der DNA auf die Bindungsaffinität DNA-bindender Proteine zu ihren Zielsequenzen

Die Proteine MBD1, MBD2, MBD3 und MeCP2 binden in Säugetieren spezifisch an methylierte DNA-Abschnitte. Als mC-Reader übersetzen sie die mit der DNA-Methylierung verbundene Information^[387] in ein transkriptionsinhibierendes Signal.^[388] Dies geschieht über die Rekrutierung von histonmodifizierenden Enzymen. So interagiert MBD1 mit der Histonmethyltransferase SETDB1,^[389] während MBD2 und MeCP2 Histondeacetylasen rekrutieren.^[390-391] Durch die Wirkung dieser Enzyme kommt es zu einer dichteren Packung des Chromatins in der Umgebung der methylierten DNA-Abschnitte, welche die Transkription in diesen Bereichen verhindert.^[392-395] Alle Proteine der MBD-Familie enthalten eine Methyl-CpG-bindende Domäne, welche spezifisch an symmetrisch methylierte CpG-Sequenzpalindrome binden kann.^[388] Nach der Entdeckung von hmC haben verschiedene Gruppen das Bindungsverhalten der MBD-Proteine gegenüber hemi- und vollständig hydroxymethylierten CpGs untersucht.^[135, 285, 396-397] Trotz z.T. abweichender Ergebnisse gibt es zunehmend Belege dafür, dass besonders MeCP2 und auch MBD3 an hydroxymethylierte CpGs binden können. So berichten Mellen *et al.*, dass MeCP2 in Mäusehirnen das wichtigste hmC-bindende Protein ist und dort eine ähnliche Bindungsaffinität gegenüber methylierten und hydroxymethylierten DNA-Abschnitten aufweist.^[285] Außerdem wurde von Yildirim *et al.* in embryonalen Stammzellen eine Bindung von MBD3 zu hmC-haltigen Sequenzabschnitten nachgewiesen.^[135] Trotzdem scheint die

Bindungsaffinität der MBD-Proteine zu hydroxymethylierten CpGs eher geringer als zu methylierten CpGs zu sein, und die Proteine MBD1 und 2 binden offenbar nur an methylierte DNA-Abschnitte.^[285, 396-398]

Die Zinkfinger/CXXC-Domäne ist das gemeinsame Element einer weiteren Gruppe von DNA-bindenden Proteinen, zu der auch die Tet-Enzyme gehören. Viele dieser Proteine mit CXXC-Domäne sind durch Modifizierung der Chromatinstruktur direkt an der Transkriptionsregulation beteiligt.^[399] So wurde gezeigt, dass die histonmodifizierenden Enzyme CFP1 (CXXC Finger Protein 1) und KDM (Lysindemethylase) für die Steuerung der Lysinmethylierung an den Histonfortsätzen verantwortlich sind.^[400-401] Dabei kann z.B. die Lysindemethylase KDM2A das Lysin 36 von Histon 3 (H3K36) selektiv demethylieren.^[402-403] Im Gegensatz zur den Mitgliedern der MBD-Familie, die bevorzugt mit methylierten CpGs interagieren, binden Proteine mit CXXC-Domäne eher an unmethylierte CpG-Inseln.^[401, 404] Es ist davon auszugehen, dass ähnlich wie bei der MBD-Familie die Bindungsaffinität von Proteinen mit CXXC-Domäne zu methylierten und hydroxymethylierten DNA-Abschnitten unterschiedlich ausfällt.

Auch die Bindung von Proteinen mit SRA-Domäne (*Set and RING finger-associated domain*) kann durch das Vorhandensein modifizierter Cytosine in der DNA beeinflusst werden.^[405] So kann das Protein UHRF1, welches essentiell für die Funktion und Strangselektivität der *maintenance*-Methyltransferase DNMT1 ist, über seine SRA-Domäne sowohl an methylierte als auch an hydroxymethylierte DNA-Stränge binden.^[397, 406] UHRF2 besitzt dagegen eine Präferenz für hydroxymethylierte Abschnitte der DNA.^[41, 407]

Wie oben beschrieben, kann eine Änderung des CpG-Substitutionsmusters, wie etwa die Oxidation von mC zu hmC, die Bindungsaffinität von DNA-bindenden Proteinen zu bestimmten Zielsequenzen entscheidend verändern. Da viele der MBD- und CXXC-Proteine die Chromatinstruktur an ihren Bindungsstellen direkt oder indirekt beeinflussen, ist auf diese Weise eine Regulation der Transkription möglich.

3.5.2 Massenspektrometriebasierte Identifikation von hmC-, fC-, und caC-bindenden Proteinen

Eine massenspektrometriebasierte Suche nach bisher unbekannten Readern modifizierter Cytosinbasen wurde von den Gruppen um Vermeulen und Carell einerseits^[41] und Reik und Balasubramanian andererseits^[296] durchgeführt.

Von Vermeulen, Carell und Mitarbeitern wurden embryonale Stammzellen (ESCs), neurale Vorläuferzellen (*neural progenitor cells*, NPCs) und adultes Hirngewebe der Maus untersucht und so unterschiedliche murine Entwicklungsstadien abgedeckt. Dabei konnten eine Vielzahl neuer Proteine

identifiziert werden, die mit hydroxymethylierten, methylierten oder unmodifizierten CpGs interagieren. Das Interaktom der verschiedenen Cytosinmodifikationen erwies sich als überraschend dynamisch, d.h. die in ESCs, NPCs und adultem Hirngewebe identifizierten Reader-Sets unterscheiden sich deutlich voneinander. Dabei sind WDR76, THY28 und die Glykosylase NEIL1 die einzigen hmC-bindenden Proteine, die in allen untersuchten Entwicklungsstadien der Maus auftreten.^[41] THY28 ist ein evolutionär stark konserviertes Protein, das sowohl in Wirbeltieren als auch in Pflanzen vorkommt.^[408-409] Es scheint einen Einfluss auf apoptotische Prozesse zu haben, ist aber bisher noch nicht genau charakterisiert worden.^[410] Auch die Funktion von WDR76 ist bisher unbekannt, das Protein kann jedoch u.a. mit OCR / SPINDLIN-1 und HELLS / LSH interagieren.^[41] HELLS ist eine murine DNA-Helikase, deren Beteiligung an der Regulation des mC-Levels diskutiert wird, während OCR ein Protein ist, das spezifisch an die Histonmodifikation H3K4me3 binden kann.^[411-412] Diese Modifikation ist charakteristisch für euchromatische Bereiche des Genoms mit aktiver Gentranskription.^[413-414]

Interessanterweise treten die meisten der von Vermeulen, Carell und Mitarbeitern identifizierten hmC-Reader nur zu bestimmten Zeitpunkten in der Entwicklung der Maus in Erscheinung. So gehört UHRF2 zu den wichtigsten hmC-Readern in neuronalen Vorläuferzellen. In diesen Zellen scheint UHRF2 die iterative Oxidationskapazität der Tet-Enzyme zu stimulieren, so dass verstärkt die höher oxidierten Basen fC und caC gebildet werden.^[41] In embryonalen Stammzellen wird UHRF2 dagegen nicht exprimiert, und auch im Hirn von adulten Tieren scheint UHRF2 keine herausragende Bedeutung zu besitzen.^[415] Hier wurden stattdessen der Transkriptionsregulator THAP11 / RONIN und der für die DNA-Replikation wesentliche Proteinkomplex RFC als wichtige hmC-Interaktoren identifiziert.^[41, 416-418] Obwohl sich die von Vermeulen, Carell und Mitarbeitern identifizierten Reader-Sets von hmC und mC größtenteils voneinander unterscheiden, wurden auch Proteine identifiziert, die mit beiden Modifikationen interagieren können. Dazu zählen u.a. MeCP2 und UHRF1, deren Fähigkeit zur Bindung von hydroxymethylierten und methylierten CpGs auch von anderen Gruppen gefunden wurde.^[41, 285, 391, 406]

Eine Suche nach fC- und caC-spezifischen Readern wurde von Vermeulen, Carell und Mitarbeitern in mESC-Extrakten durchgeführt.^[41] Viele der dabei identifizierten Proteine spielen bei der DNA-Reparatur eine Rolle. Außerdem wurde die Glykosylase TDG, die von entscheidender Bedeutung für die BER-basierte aktive DNA-Demethylierung ist (siehe Abschnitt 3.4.2.1), auch in dieser Studie als fC- und caC-Interaktor identifiziert.^[34] Interessanterweise konnte daneben eine spezifische Interaktion zwischen DNMT1 und caC nachgewiesen werden.

Auch in der von Reik, Balasubramanian und Mitarbeitern durchgeführten Studie konnten neue Reader von mC, hmC und fC in murinen Stammzellextrakten identifiziert werden.^[296] Dabei zeigte sich, dass

die Spezifität der mit den modifizierten Cytosinbasen interagierenden Proteine nicht nur vom Basentyp, sondern auch von der DNA-Sequenz abhängt.

Überraschenderweise wurden von Reik, Balasubramanian und Mitarbeitern nur wenige Proteine gefunden, die eine hohe Präferenz für die Bindung an hydroxymethylierte DNA-Abschnitte aufweisen. Zu diesen gehören u.a. der Coaktivator RBM14 / CoAA und das Protein RPL26, welches die Expression des Transkriptionsfaktors p53 kontrolliert.^[419-421] Die Anzahl der in diesem Screen identifizierten fC-Reader ist dagegen deutlich größer, wobei viele dieser Proteine an der Regulation der Transkription oder der Organisation der Chromatinstruktur beteiligt sind. So sind mehrere Untereinheiten des NuRD-Komplexes in der Gruppe der spezifisch mit fC interagierenden Proteine angereichert. Der NuRD-Komplex kann mit Hilfe seiner Histondeacetylaseuntereinheiten HDAC1 und HDAC2 die Transkription inhibieren.^[422-423] Es wurde gezeigt, dass NuRD durch Interaktion mit dem Methyl-CpG-bindendem Protein MBD2 zu methylierten DNA-Abschnitten dirigiert wird.^[424] NuRD ist somit ein Vermittler zwischen der im DNA-Methylierungsmuster codierten epigenetischen Information, der Histondeacetylierung und dem damit verbundenem Gen-Silencing. Die Ergebnisse von Reik, Balasubramanian und Mitarbeitern deuten nun darauf hin, dass der NuRD-Komplex direkt oder indirekt auch von fC-haltigen DNA-Abschnitten rekrutiert werden kann.

Weiterhin wurden von Reik, Balasubramanian und Mitarbeitern mehrere Forkhead-Box-Proteine (FOX-Proteine) als potentielle fC-Reader identifiziert. Die FOX-Proteine sind eine Familie von eukaryotischen Transkriptionsfaktoren, welche über ihre Forkhead- oder Winged-Helix-Domäne an die DNA binden und direkt die lokale Chromatinstruktur beeinflussen können.^[425-426] Sie steuern auf diese Weise eine Vielzahl von biologischen Prozessen, u.a. als terminale Effektoren in wichtigen zellinternen Signalwegen.^[427] Die von Reik, Balasubramanian und Mitarbeitern gefundene spezifische Interaktion zwischen den FOX-Proteinen FOXK2, FOXI3, FOXP1, FOXP4 und fC ist ein weiterer Hinweis auf die eigenständige epigenetische Bedeutung dieser Cytosinmodifikation.

Bei den beiden oben beschriebenen Screenings wurde eine ganze Reihe von Proteinen identifiziert, die spezifisch an die Cytosinmodifikationen hmC, fC und caC binden können. Einige dieser Proteine können z.B. als chromatinmodifizierende Enzyme direkt oder indirekt die Transkription beeinflussen. Es ist daher wahrscheinlich, dass besonders hmC, aber möglicherweise auch fC, neben ihrer Rolle bei der aktiven DNA-Demethylierung eine Bedeutung als eigenständige epigenetische Basen besitzen und ähnlich wie mC einen bestimmten Transkriptionszustand codieren können.^[40, 297]

4 Ergebnisse und Publikationen

4.1 “Parallel Isotope-Based Quantification of Modified tRNA Nucleosides”

T. Brückl, D. Globisch, **M. Wagner**, M. Müller, T. Carell, *Angew. Chem. Int. Ed. Engl.* **2009**, *48*, 7932-7934.

Das systemische Zusammenspiel der modifizierten Nukleoside einer tRNA oder eines tRNA-Sets als Gesamtheit wurde lange Zeit kaum untersucht. Ein Grund hierfür war das Fehlen einer ausgereiften Methode zur parallelen Quantifizierung von modifizierten RNA-Nukleosiden. Ziel dieser Arbeit war es daher, eine solche Quantifizierungsmethode zu entwickeln.

Die hier vorgestellte LC-MS-basierte Methode erlaubt als Isotopenverdünnungsmethode eine exakte Quantifizierung der absoluten Nukleosidmengen in der Probe. Sie ist außerdem im Vergleich zu bis dato üblichen UV-basierten Quantifizierungsverfahren um ein Vielfaches sensitiver. Ihre herausragende Leistungsfähigkeit wird mit einer Analyse der tRNA-Modifikationen in humanen Krebszelllinien, E.coli und Schweineleber demonstriert.

Für diese Arbeit habe ich zwei humane Krebszelllinien in Kultur genommen. Aus den Zellen habe ich die tRNA isoliert und diese enzymatisch auf das Nukleosidlevel verdaut. Anschließend habe ich die Proben mit der hier vorgestellten, von Tobias Brückl und Daniel Globisch entwickelten Methode analysiert und dabei sechs in der tRNA vorkommende Modifikationen quantifiziert.

Parallel Isotope-Based Quantification of Modified tRNA Nucleosides**

Tobias Brückl, Daniel Globisch, Mirko Wagner, Markus Müller, and Thomas Carell*

Transfer RNAs (tRNAs) are the essential adapter molecules needed for the translation of the genetic code into a peptide sequence at the ribosome. Human cells possess 49 different isoacceptor tRNAs, each carrying a specific amino acid at the 3' terminus.^[1] All tRNA molecules consist of 70–100 ribonucleotides of which up to 20 % are modified. The large number (>95) of different modifications known today^[2,3] and in particular their structural diversity establish a second layer of molecular information most likely required for the decoding process.^[4] However, only limited information is available as to how these modifications influence the translation of genetic information,^[5] the folding of tRNA,^[4,6] and tRNA function.^[7]

In the current era of “omics” research the central scientific approach is the study of cellular components not individually but context based.^[8] In our opinion, a new quantitative method is also needed for the analysis of modified tRNA nucleosides.^[3,9] As the literature indicates that tumor cells possess a significantly different tRNA modification pattern than that in normal cells,^[10,11] particularly in their mitochondria,^[12] such a method would lead to a better characterization of tumor cells and may enable deeper insight in the metabolism of tumor cells.^[11] Here, we report an isotope-based, mass-spectrometric method that allows direct, parallel quantification of in principle all tRNA modifications with extremely high sensitivity. We provide initial data showing that surprisingly large variations are observed among tumor cells and also between tumor cells and non-tumorigenic tissue.

The method is represented in Figure 1. Whole tRNA is isolated from cells or tissue using a modified literature protocol.^[13] The bulk tRNA material is subsequently fully digested to obtain a mixture of the tRNA nucleosides.^[14] In order to quantify certain modifications, we determined calibration curves for each modification individually and added precisely known amounts of synthetic isotope-labeled, modified tRNA nucleosides to the digests. To develop the method we first attempted to quantify the tRNA modifications m^6A , m^2A , Am , t^6A , i^6A , and ms^2i^6A ,^[15] because these

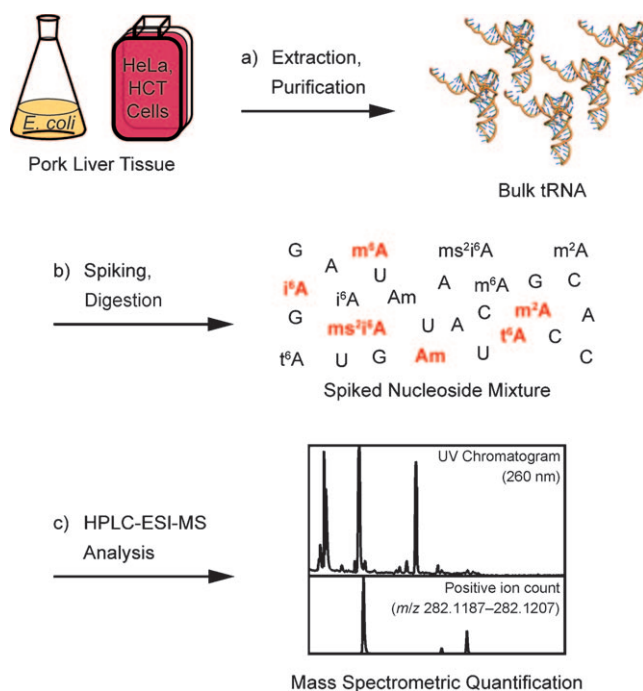


Figure 1. Depiction of the quantitative method for analysis of modified tRNA nucleosides. a) Extraction: phenol extraction; purification: PD10 chromatography, anion-exchange chromatography; b) Spiking: addition of isotope-labeled standards; Digestion: enzymatic digestion of the bulk tRNA mixture; black symbols show natural nucleosides; red symbols indicate added isotope-labeled nucleosides; c) HPLC-ESI-MS analysis: chromatographic separation and mass spectrometric analysis (for experimental details see the Supporting Information).

nucleosides are mostly present directly 3'-adjacent to the anticodon^[2,3] and therefore likely to be directly involved in the decoding process.^[16] The nucleoside ms^2i^6A is a particularly interesting modification because it exists only in prokaryotes and mitochondrial tRNAs.^[2,3] As such, study of this nucleoside may allow insight into the mitochondrial activity of cells. All mentioned modified nucleosides were synthesized as isotope-labeled (*) derivatives as indicated in Figure 2.^[17] The digest mixture containing the added isotope-labeled compounds was subsequently analyzed by HPLC-ESI-MS.

Typical UV traces of the obtained *E. coli* and pork liver modification pattern are depicted in Figure 3a and 3b, respectively. Next to the dominant signals for the canonical RNA nucleosides, the various modified nucleosides either appear as small signals or are hidden in the base line. The modifications were assigned unambiguously by MS analysis. In Figure 3c the ion counts of m^2A and m^6A in comparison to

[*] T. Brückl,^[1] D. Globisch,^[1] M. Wagner, Dr. M. Müller, Prof. Dr. T. Carell
Center for Integrated Protein Science (CiPSM)
Department of Chemistry and Biochemistry, LMU Munich
Butenandtstrasse 5–13, 81377 Munich (Germany)
Fax: (+49) 89-2180-77756
E-mail: thomas.carell@cup.uni-muenchen.de

[†] These authors contributed equally to this work.

[**] We thank the CiPSM Cluster, SFB 749, the Fonds of the German Chemical Industry, and Bayer-Schering for financial support.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.200902740>.

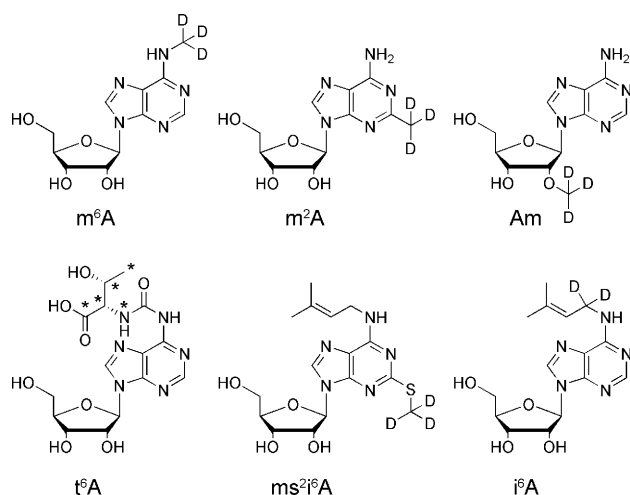


Figure 2. Heavy-atom-labeled adenosine modifications investigated. Asterisks (*) are used to label ^{15}N and ^{13}C atoms; D = deuterium. Base-modified nucleosides occur 3'-adjacent to the anticodon in position 37. The ribose-modified nucleoside Am is present in position 4.

the added synthetic isotope-labeled species are depicted as an example. Because of the mass difference they appear in the mass spectra as two separate signals allowing quantification by integration of the area under the mass peaks. The high quality of the obtained data with an average error margin of only 5% proves that the isotope dilution method can be readily applied for the analysis of tRNA modifications if isotope-labeled, modified RNA nucleosides are available.^[18]

With the new method for tRNA analysis in hand we started to investigate how the modification patterns of tumor cells differ from each other and from healthy tissue. For an initial screening we used the human epithelial cell lines HCT-116 and HeLa, which were derived from a colorectal tumor and from an adenocarcinoma of the cervix, respectively. For comparison, we also analyzed the modification pattern of pork liver and *E. coli*. The results of the determinations are depicted in Figure 4. In agreement with literature data we detect the modification m^2A only in *E. coli* and not in eukaryotic cells. In contrast, the Am modification^[19] is found only in eukaryotic cells.^[2,3,20] To our surprise we observed that the level of t^6A in healthy pork liver is significantly lower than that in both tumor cell lines, which indicates that the biosynthesis of this modification may be upregulated in HeLa and HCT-116 cells. Further significant differences between the tumor cell lines were noted for m^6A and i^6A . HeLa cells contain twice as much m^6A than HCT-116 cells, which exhibit the same amount of m^6A as pork liver tissue. The level of i^6A is 30% lower in HeLa cells than in HCT-116 cells, whereas HeLa cells contain the same amount of i^6A as healthy liver tissue. These results support the idea that the analysis of modified tRNA nucleosides can differentiate between tumorigenic and non-tumorigenic cells and that even discrimination of different cancer cell lines may become possible once the analysis is extended to a broader variety of nucleoside modifications.

We next addressed the question of how the mitochondrial tRNA modification pattern changes in tumor cells. It has been

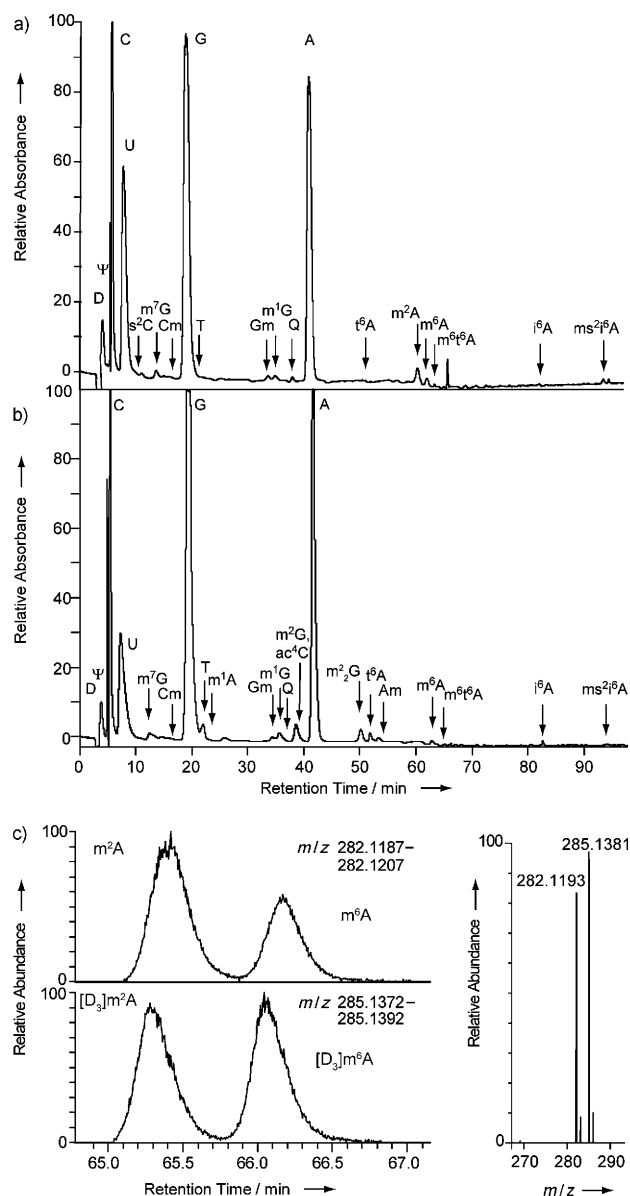


Figure 3. a) HPLC trace of the tRNA digest from *E. coli* ($\lambda = 260$ nm). b) HPLC trace of the tRNA digest from pork liver ($\lambda = 260$ nm). c) Positive ion traces of the protonated nucleosides m^2A and m^6A and the corresponding synthetic isotope-labeled derivatives (left). Relevant high-resolution mass spectrometric data for unlabeled and labeled m^2A (right).

known for some time that tumors have a reduced oxidative phosphorylation activity and that many tumor cells derive most of their needed energy from glycolysis,^[21,22] which induces a reduced pH value in tumor tissue.^[21] In order to investigate this phenomenon we used the isotope method for quantification of ms^2i^6A , which is present only in mitochondria of eukaryotic cells.

The result of this analysis is depicted in Figure 4 (also see inset). As expected, the modification ms^2i^6A is present in the bacterium *E. coli*. In addition, it is detectable in small but significant amounts in pork liver. The ability to quantify mitochondrial ms^2i^6A proves again the high sensitivity of the isotope-based method.

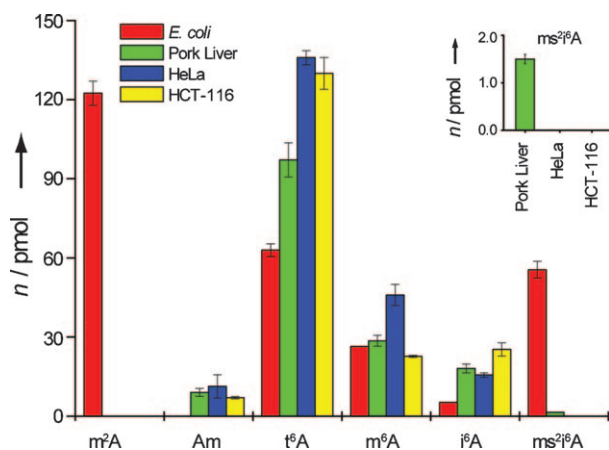


Figure 4. Comparison of modified nucleoside levels of *E. coli*, pork liver, HeLa, and HCT-116 cell lines. Error bars represent the standard deviations calculated from multiple experiments.

To our surprise we found that the analyzed tumor cells do not contain the modification ms²t⁶A at all, indicating that the mitochondrial tRNAs in tumor cells and normal cells indeed differ substantially. This result provides a first indication that the impairment of mitochondria in tumor cells (Warburg hypothesis)^[23] is detectable using our isotope-based analysis method.

Even though our interpretation is currently based on the analysis of only six tRNA modifications, the results clearly show that the reported isotope-labeling method can rapidly provide quantitative data about tRNA modification levels in cells and tissues. Even quantification of modifications present only in mitochondrial tRNA is possible. Assuming that different species, various mammalian tissues, and different tumor tissues possess substantially different patterns of modified tRNA nucleosides,^[24] then this isotope method opens new doors for the detailed analysis of cells and tissues. The presented method, for which we suggest the name “modiomics”, might pave the way for the detailed analysis of the function and biosynthesis of modified tRNA nucleosides.

Received: May 22, 2009

Published online: September 11, 2009

Keywords: bioanalytical methods · gene translation · isotope labeling · modiomics · tRNA

- [1] J. M. Goodenbour, T. Pan, *Nucleic Acids Res.* **2006**, *34*, 6137–6146.
- [2] F. Juehling, M. Moerl, R. K. Hartmann, M. Sprinzl, P. F. Stadler, J. Puetz, *Nucleic Acids Res.* **2009**, *37*, D159–D162; J. Rozenski, P. F. Crain, J. A. McCloskey, *Nucleic Acids Res.* **1999**, *27*, 196–197; H. Grosjean, M. Sprinzl, S. Steinberg, *Biochimie* **1995**, *77*, 139–141.
- [3] A. Czerwonec, S. Dunin-Horkawicz, E. Purta, K. H. Kaminska, J. M. Kasprzak, J. M. Bujnicki, H. Grosjean, K. Rother, *Nucleic Acids Res.* **2009**, *37*, D118–D121.
- [4] M. Helm, *Nucleic Acids Res.* **2006**, *34*, 721–733.
- [5] G. R. Bjork, J. M. B. Durand, T. G. Hagervall, R. Leipunviene, H. K. Lundgren, K. Nilsson, P. Chen, Q. Qian, J. Urbonavicius,

- FEBS Lett.* **1999**, *452*, 47–51; E. M. Gustilo, F. A. P. Vendeix, P. F. Agris, *Curr. Opin. Microbiol.* **2008**, *11*, 134–140; W. F. Waas, Z. Druzina, M. Hanan, P. Schimmel, *J. Biol. Chem.* **2007**, *282*, 26026–26034.
- [6] L. Randau, B. J. Stanley, A. Kohlway, S. Mechta, Y. Xiong, D. Soell, *Science* **2009**, *324*, 657–659.
- [7] U. Z. Littauer, H. Inouye, *Annu. Rev. Biochem.* **1973**, *42*, 439–470; B. C. Persson, *Mol. Microbiol.* **1993**, *8*, 1011–1016; M. Buck, B. N. Ames, *Cell* **1984**, *36*, 523–531.
- [8] R. Aebersold, M. Mann, *Nature* **2003**, *422*, 198–207; B. F. Cravatt, G. M. Simon, J. R. Yates III, *Nature* **2007**, *450*, 991–1000; L. Assogba, A. Ahamada-Himidi, N. M. B. Habich, D. Aoun, L. Boukli, F. Massicot, C. M. Mounier, J. Huet, A. Lamouri, J. E. Ombetta, J. J. Godfroid, C. Z. Dong, F. Heymans, *Eur. J. Med. Chem.* **2005**, *40*, 850–861; S. Broder, J. C. Venter, *Curr. Opin. Biotechnol.* **2000**, *11*, 581–585; D. J. Lockhart, E. A. Winzler, *Nature* **2000**, *405*, 827–836.
- [9] H. Grosjean, C. Gaspin, C. Marck, A. Decatur Wayne, V. de Crecy-Lagard, *BMC Genomics* **2008**, *9*, 470.
- [10] G. Dirheimer, W. Baranowski, G. Keith, *Biochimie* **1995**, *77*, 99–103.
- [11] A. Frickenschmidt, H. Froehlich, D. Bullinger, A. Zell, S. Laufer, C. H. Gleiter, H. Liebich, B. Kammerer, *Biomarkers* **2008**, *13*, 435–449.
- [12] K. Watanabe, *Bull. Chem. Soc. Jpn.* **2007**, *80*, 1253–1267.
- [13] M. Buck, M. Connick, B. N. Ames, *Anal. Biochem.* **1983**, *129*, 1–13.
- [14] P. F. Crain, *Methods Enzymol.* **1990**, *193*, 782–790.
- [15] Abbreviations: m⁶A: N⁶-methyladenosine, m²A: 2-methyladenosine, Am: 2'-O-methyladenosine, t⁶A: N⁶-threonylcarbamoyl-adenosine, ms²t⁶A: 2-methylthio-N⁶-isopentenyladenosine, i⁶A: N⁶-isopentenyladenosine.
- [16] F. V. Murphy, V. Ramakrishnan, A. Malkiewicz, P. F. Agris, *Nat. Struct. Mol. Biol.* **2004**, *11*, 1186–1191; J. M. B. Durand, G. R. Bjork, A. Kuwae, M. Yoshikawa, C. Sasakawa, *J. Bacteriol.* **1997**, *179*, 5777–5782.
- [17] The syntheses of isotope-labeled compounds will be reported elsewhere. Examples for syntheses of tRNA nucleosides: C. Hoebartner, C. Kreutz, E. Flecker, E. Ottenschlaeger, W. Pils, K. Grubmayr, R. Micura, *Monatsh. Chem.* **2003**, *134*, 851–873; M. Sundaram, P. F. Crain, D. R. Davis, *J. Org. Chem.* **2000**, *65*, 5609–5614; R. M. Hyde, A. D. Broom, R. W. Buckheit, Jr., *J. Med. Chem.* **2003**, *46*, 1878–1885; K. Hirota, Y. Kitade, Y. Kanbe, Y. Maki, *J. Org. Chem.* **1992**, *57*, 5268–5270; M. H. Fleyscher, *J. Med. Chem.* **1972**, *15*, 187–191; E. Kierzek, R. Kierzek, *Nucleic Acids Res.* **2003**, *31*, 4461–4471; F. Klepper, E.-M. Jahn, V. Hickmann, T. Carell, *Angew. Chem.* **2007**, *119*, 2377–2379; *Angew. Chem. Int. Ed.* **2007**, *46*, 2325–2327; T. Brückl, F. Klepper, K. Gutsmedl, T. Carell, *Org. Biomol. Chem.* **2007**, *5*, 3821–3825.
- [18] J. J. Dalluge, T. Hashizume, J. A. McCloskey, *Nucleic Acids Res.* **1996**, *24*, 3242–3245. For isotope-labeled tRNA nucleosides please contact Thomas Carell.
- [19] M. L. Wilkinson, S. M. Crary, J. E. Jackman, E. J. Grayhack, E. M. Phizicky, *RNA* **2007**, *13*, 404–413.
- [20] C. W. Gehrke, K. C. Kuo, *J. Chromatogr.* **1989**, *471*, 3–36.
- [21] R. A. Gatenby, R. J. Gillies, *Nat. Rev. Cancer* **2004**, *4*, 891–899.
- [22] I. Samudio, M. Fiegl, M. Andreeff, *Cancer Res.* **2009**, *69*, 2163–2166; C. Frezza, E. Gottlieb, *Semin. Cancer Biol.* **2009**, *19*, 4–11; V. Gogvadze, S. Orrenius, B. Zhivotovsky, *Trends Cell Biol.* **2008**, *18*, 165–173.
- [23] M. A. Kiebish, X. Han, H. Cheng, J. H. Chuang, T. N. Seyfried, *J. Lipid Res.* **2008**, *49*, 2545–2556.
- [24] M. W. Taylor, G. A. Granger, C. A. Buck, J. J. Holland, *Proc. Natl. Acad. Sci. USA* **1967**, *57*, 1712–1719; A. Costa, J.-P. Pais de Barros, G. Keith, W. Baranowski, J. Desgres, *J. Chromatogr. B* **2004**, *801*, 237–247.

4.2 “Isotope-Based Analysis of Modified tRNA Nucleosides Correlates Modification Density with translational Efficiency”

C. Brandmayr[#], **M. Wagner[#]**, T. Brückl, D. Globisch, D. Pearson, A. C. Kneuttinger, V. Reiter, A. Hienzsch, S. Koch, I. Thoma, P. Thumbs, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angew. Chem. Int. Ed. Engl.* **2012**, 51, 11162-11165.

Co-first authors

Ziel dieses Projektes war es, die tRNA-Modifikationen in den Geweben und Organen von Säugetieren zu analysieren. Mit der von Tobias Brückl und Daniel Globisch entwickelten LC-MS-basierten Quantifizierungsmethode (siehe 4.1) sollte dazu in den untersuchten Organen eine repräsentative Auswahl an tRNA-Modifikationen parallel quantifiziert werden. Damit sollte der wissenschaftlichen Gemeinschaft ein Datensatz zur Verfügung gestellt werden, der die Häufigkeit und Verteilung dieser Modifikationen in den verschiedenen Organen erstmals exakt wiedergibt.

Es konnte in dieser Publikation gezeigt werden, dass die tRNA-Sets in Säugetieren abhängig vom Organ eine unterschiedliche und für das jeweilige Organ charakteristische Modifikationsdichte aufweisen. Die Modifikationsdichte des tRNA-Sets korreliert dabei positiv mit dem Proteinsynthesebedarf des Organs. Diese Ergebnisse weisen darauf hin, dass in Säugetieren über den Modifikationsgrad des tRNA-Sets eine Regulation der Proteinsyntheserate möglich ist.

Die tRNA-Sets der Schweineorgane waren bereits von Tobias Brückl während seiner Doktorarbeit untersucht worden. Er hat auch den Zusammenhang zwischen der organspezifischen Modifikationsdichte und dem Proteinsynthesebedarf der Organe erkannt. In Zusammenarbeit mit Caterina Brandmayr habe ich die tRNA-Sets der Maus untersucht. Dabei war ich zuständig für die Extraktion der tRNA, die Quantifizierung der Modifikationen und die Datenanalyse. Ich war auch an der Anfertigung des Manuskripts und der darin enthaltenen Grafiken beteiligt.

Isotope-Based Analysis of Modified tRNA Nucleosides Correlates Modification Density with Translational Efficiency**

Caterina Brandmayr, Mirko Wagner, Tobias Brückl, Daniel Globisch, David Pearson, Andrea Christa Kneutinger, Veronika Reiter, Antje Hienzsch, Susanne Koch, Ines Thoma, Peter Thumbs, Stylianos Michalakis, Markus Müller, Martin Biel, and Thomas Carell*

Transfer RNAs (tRNAs) are adapter molecules needed to translate genetic information into a peptide sequence.^[1] At the ribosome, the anticodon of each tRNA reads the corresponding codon of the messenger RNA. This anticodon–codon interaction allows the ribosome's large subunit to catalyze amide-bond formation between the cognate amino acids present at the 3' terminus of aminoacyl-tRNAs and the growing peptide chain.^[2] The tRNA adapters required for this process display a surprisingly large chemical diversity.^[3] Aside from the four canonical nucleosides A, C, G, and U, more than 100 modified nucleosides are key constituents (Figure 1).^[4] The most diverse and complex chemical structures are found in the anticodon stem-loop either in the anticodon at the wobble position or directly adjacent to the 3' position of the anticodon,^[5] suggesting that here the chemical complexity is necessary to establish translational fidelity.^[6] The ribosome seems to need the modified anticodon region to better distinguish correctly base-paired tRNA from mispaired interactions in order to prohibit, for example, codon-slippage processes that would lead to frameshifts.^[7]

In order to investigate how the set of nucleoside modifications influences the translational efficiency we quantified the tRNA modifications individually in various tissues by an isotope-dilution-based LC–MS method. (Details on the

materials and methods are given in the Supporting Information, Table S1, and Figure S1). The quantified modification levels were correlated with the translational efficiency by means of an in vitro translation system. For the experiments, 11 representative tRNA modifications (Figure 1) were chemically synthesized as isotope-labeled derivatives.^[8] A majority of the investigated nucleosides are located inside the extended anticodon,^[9] the other synthesized modified nucleosides are found at various other positions.^[10]

As biological material for the analysis we chose a range of different organ tissues from mouse and pig. Porcine tissues were used because they are available in large amounts, while murine tissues were analyzed at a later stage to confirm the results in a genetically more defined organism. For pig, 5–10 g of tissue from two animals was sampled from each organ, while murine samples were obtained from two sets of five animals of which whole organs were analyzed. After total tRNA extraction and complete enzymatic hydrolysis to nucleosides, a mixture of the isotope-labeled tRNA modifications was added and the solution was subjected to LC–MS analysis. The ratios of the mass peak integrals from natural to isotope-labeled nucleosides were determined and calibration curves, which were previously measured for each investigated modified nucleoside, then allowed exact parallel quantification of the respective modifications (see Figure S2 in the Supporting Information).^[8a] LC–MS quantification was performed at least in triplicate and results were averaged for each tissue. The error margin of the experiment was in this way limited to around 5%.

The obtained quantitative values for mouse and pig samples are shown color coded in Figure 2A and B, respectively, together with the approximate positions of the measured modifications in the tRNA sequence. The values represent the measured number of each modification per 1000 tRNA molecules (‰) (exact values are listed in Tables S2–7 in the Supporting Information). Therefore, rather than yielding the absolute concentration of a modification in a given tissue, the data show directly the extent to which the analyzed tRNA set is modified. For representative murine and porcine tissues, an additional quantification of the ubiquitous tRNA modifications m⁵C and Ψ was performed (see Figure S4).

The data show that each tissue type incorporates different amounts of a specific modified nucleoside into the respective tRNA ensemble. While the tRNAs in liver tissue contain large numbers of modified nucleosides (colored red), those isolated from lung and kidney tissue feature far fewer

[*] M. Chem. C. Brandmayr,^[†] Dipl.-Chem. M. Wagner,^[†] Dr. T. Brückl,^[†] Dr. D. Globisch, Dr. D. Pearson, M. Sc. A. C. Kneutinger, Dipl.-Chem. V. Reiter, Dr. A. Hienzsch, M. Sc. I. Thoma, Dipl.-Chem. P. Thumbs, Dr. M. Müller, Prof. Dr. T. Carell
Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität München
Butenandtstrasse 5–13, 81377 Munich (Germany)
E-mail: thomas.carell@cup.lmu.de
Homepage: <http://www.carellgroup.de>

Dipl.-Biol. S. Koch, Dr. S. Michalakis, Prof. Dr. M. Biel
Center for Integrated Protein Science at the Department of Pharmacy, Ludwig-Maximilians-Universität München
Butenandtstrasse 5–13, 81377 Munich (Germany)

[†] These authors contributed equally to this work.

[**] We thank the excellence cluster CIPSM and the SFBs 646 and 749 for generous support. Further support from the Fonds der Chemischen Industrie and Bayer Pharma AG is gratefully acknowledged. C.B. is grateful to the Boehringer Ingelheim Fonds for a predoctoral fellowship.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201203769>.

Re-use of this article is permitted in accordance with the Terms and Conditions set out at <http://angewandte.org/open>.

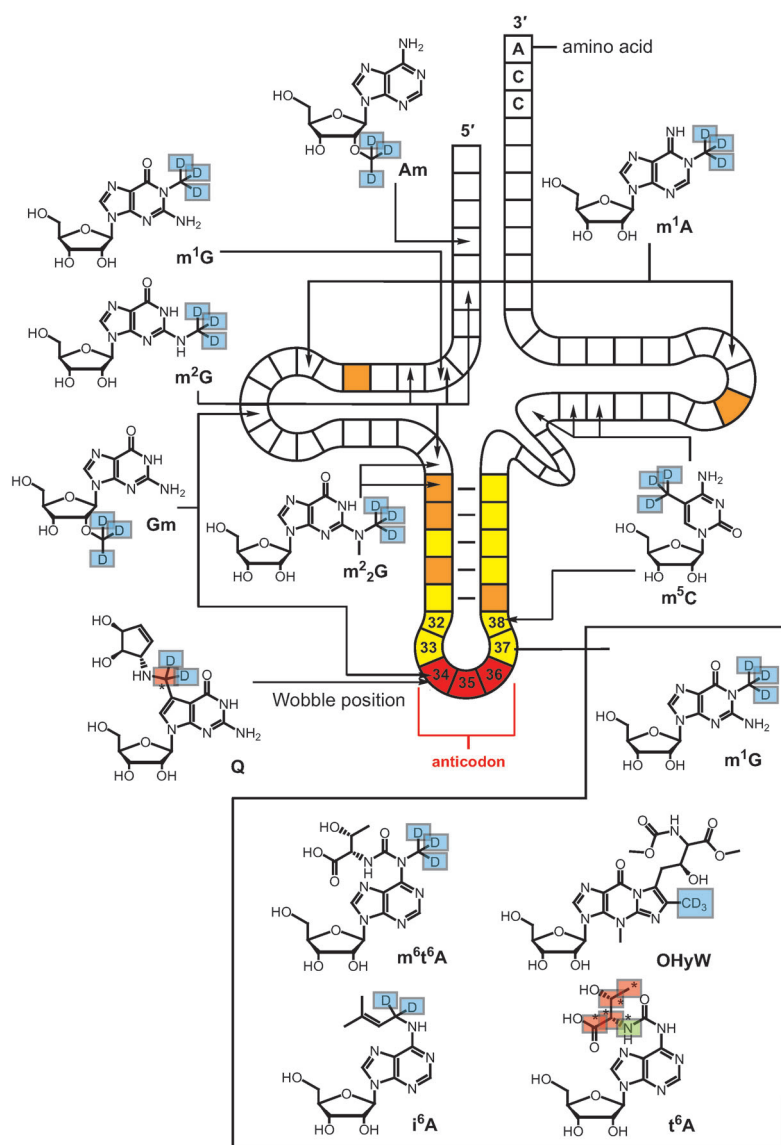


Figure 1. Isotope-labeled tRNA nucleosides present in eukaryotic tRNA and positions where these modifications are typically found. The introduced isotope labels are marked in color: D in a blue box: deuterium; * in a red box: ^{13}C ; * in a green box: ^{15}N . The anticodon is highlighted in red, the remainder of the anticodon stem-loop is in yellow, and positions of Ψ are marked in orange. Abbreviations are explained in Table S1 in the Supporting Information.

modifications (colored yellow and green). It is known that the amounts of individual tRNA species vary between tissues,^[11] but these variations are small in comparison to the detected changes in modification levels, arguing that tissues modify their tRNA to different extents (for a detailed discussion about the influence of codon bias see the Supporting Information). Most important in this respect is the observation that the levels of the m^1A , m^2G , m^5C , and Ψ modifications (Figure 2 and Figure S4) follow the overall trend. These modifications are present in almost all tRNA species, and hence if tRNA composition would bias the quantitative data then the levels of these modifications would be expected to stay constant (or at least not follow the trend strongly).

Consistently, murine and porcine tissues show similar trends with liver, characterized by a high metabolic activity, having tRNA in both cases highly modified, while the tRNA from muscle tissue such as heart shows a rather low modification content.^[12] Divergence between the two organisms is observed for some tissues such as cerebrum and spleen. This might be due to species-specific variation in tissue metabolism, or it might arise from genetic variation between species, as previously observed for bacteria.^[13] Surprising is the observation that the data for Am follow a different trend, with higher Am levels found in tissues with largely unmodified tRNA (see Figure 2 and Table S8 in the Supporting Information). Because 2'-O-methylation stabilizes RNA,^[4,14] the observed pattern could reflect the role of this modification in stabilizing hypomodified tRNA. Furthermore, while levels of queuosine (Q) generally fit the overall trend, this nucleoside has unexpectedly high levels in brain tissue both in mouse and pig, suggesting a more complex specialized role in those tissues.^[15]

In order to confirm the results, we next measured the modification content in a sequence context. To this end we carried out a parallel LC-MS analysis of partial tRNA digests (RNase A) from two representative porcine tissues, liver and heart. From the digests we obtained a number of defined tRNA fragments (small oligomers) resulting from selective cleavage after C and U. Out of the obtained fragments we determined 10 for which we were able to detect the unmodified and modified sequences using mammalian tRNA sequences from the Sprinzl tRNA database.^[3b] We then determined the relative amounts of the modified versus the unmodified tRNA fragments. The extent of modification of the respective tRNA sequence was calculated directly from the ratio between the areas of the specific mass peaks for the modified and the unmodified fragments (see Figure 3 and Table S9 in the Supporting Information).^[16] The results show that the representative modified nucleosides m^1G , m^1A , m^2G , i^6A , and t^6A are indeed more abundant in tRNA fragments derived from liver, supporting the data from the isotope-dilution-based direct nucleoside quantification.

Based on the data we concluded that tissues mature their tRNA differently to satisfy individual translational needs. In support of this hypothesis we observed that published data for the rates of protein synthesis in vivo in different mammalian organs show a good level of correlation with our quantitative data for pig tissues (Figure S5 in the Supporting Information),^[12,17] suggesting that higher overall tRNA modification content might be linked to faster rates of protein translation in a certain tissue. In order to test this hypothesis directly, we analyzed the translational efficiency of tissue-specific tRNA

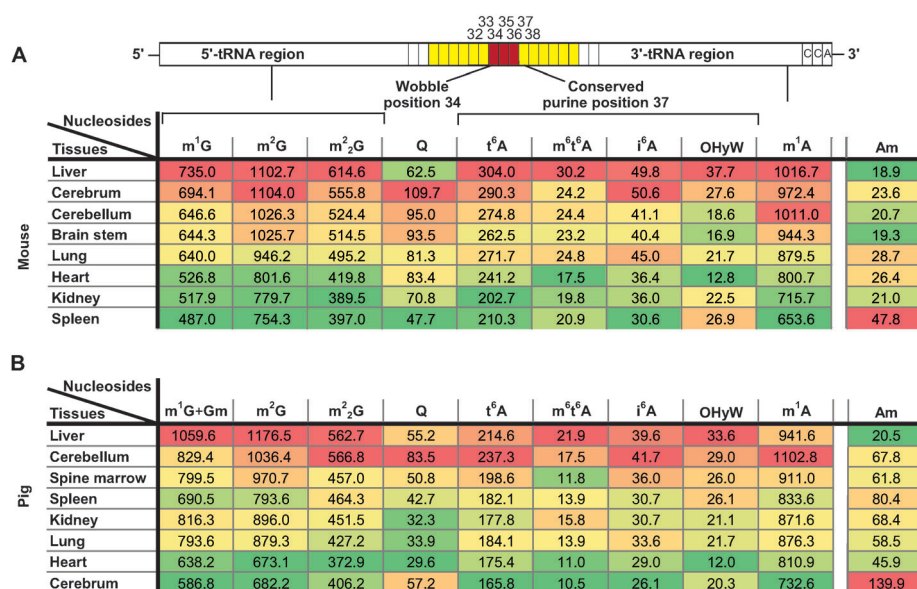


Figure 2. Quantitative data for the investigated tRNA modifications in various murine (A) and porcine (B) tissues. All tRNA nucleoside values are given per 1000 tRNA molecules (%). These data reveal a similar, tissue-dependent extent of modification for all investigated modified nucleosides except Am. Color codes in (A) and (B) are based on quantile calculations; red: highest value, yellow: 50 % quantile, green: lowest value. For intermediate values appropriate intermediate shades were calculated. Despite the slight variation in the absolute quantification values, trends in modification content were conserved across different biological samples (both in mouse and pig, see Tables S2–S5 in the Supporting Information).

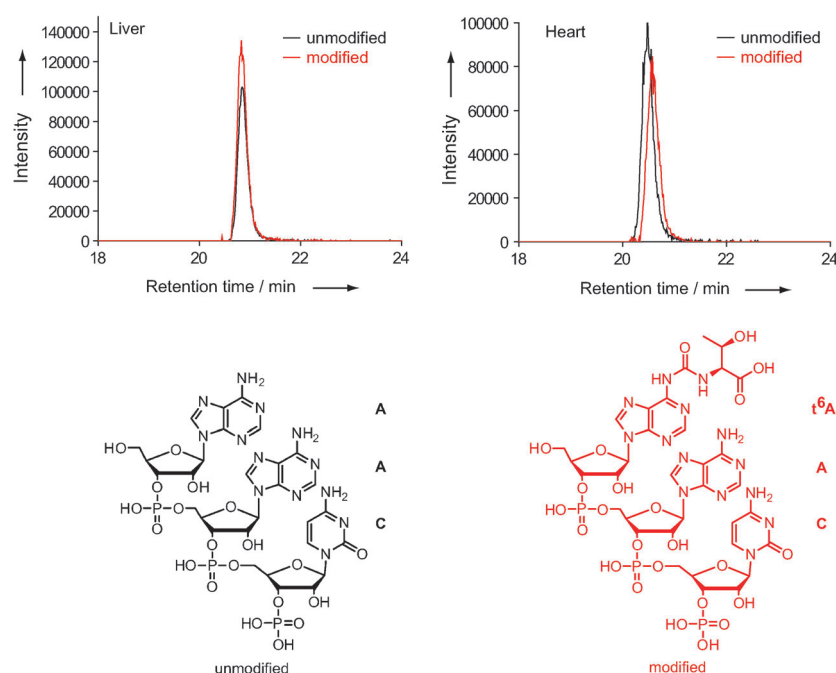


Figure 3. Representative qualitative comparison of amounts of unmodified RNA fragments AAC and the corresponding modified t⁶AAC in the RNase A digests of liver and heart tRNA. Overlaid LC–MS chromatograms showing ions detected at the calculated masses of the AAC ($m/z = 489.5682$ – 489.5742) and t⁶AAC ($m/z = 562.0863$ – 562.0933) fragments ($z = -2$) and the corresponding structures. The ratio of the peak areas of modified to unmodified fragments for liver can be seen to be higher than that for heart. Further identified fragment ratios are listed in Table S9 in the Supporting Information.

ensembles using an in vitro coupled transcription/translation reticulocyte lysate system.^[18] The original tRNAs present in the system were removed chromatographically using an ethanolamine–Sephacrose column.^[18] Subsequently, the tissue-extracted tRNA ensembles were added. Translational efficiency was measured by observing the increase in luminescence linked to the production of the protein luciferase (see Figure S7). The slopes of the plotted curves from at least three repeated experiments were normalized to the most efficient ensemble.

In a first set of experiments, total tRNA ensembles from porcine tissues were used. Figure 4 plots the measured rates against the corresponding normalized modification levels calculated based on LC–MS data presented in Figure 2B (exact values are listed in Figure S8 in the Supporting Information). From Figure 4 it can be seen that the overall modification content correlates with the translational efficiency of the isolated tRNA ensemble, but the correlation is far from optimal.

We noted that specifically the values obtained from tissues known to be rich in mitochondria (heart in particular) deviate from the expected trend.^[19] Since mitochondrial tRNA features its own set of modifications,^[13] we therefore removed the mitochondria from the porcine tissues before the tRNA extraction (Figure S6 in the Supporting Information).^[20] The obtained data for cytosolic tRNA show indeed a higher degree of correlation between the translational activity and the modification content (black squares in Figure 4; $r = 0.861$, $P = 0.028$), indicating that the modification level of cytosolic tRNAs is one factor that influences the efficiency of translation. This correlation was further confirmed for mouse using total tRNA ensembles extracted from tissues known to have relatively low mitochondrial tRNA content (see Figure S9). Our results are

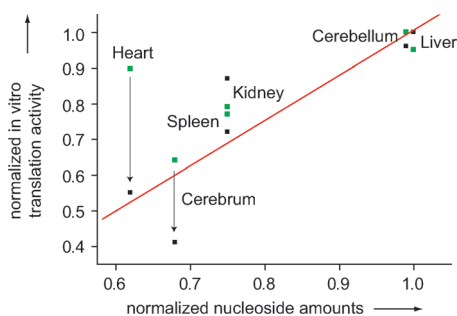


Figure 4. Translation activity of tRNA extracts isolated from different porcine tissues with (black squares) and without (green squares) removal of mitochondrial tRNAs. A plot of the linear fit of relative in vitro translation activity and normalized nucleoside levels shows a significant correlation after removal of mitochondria (red line; $r=0.861$, $P=0.028$).

consequently in good agreement with the common idea that specific noncanonical bases fine-tune the binding of tRNAs to the ribosome. As the translation rate is determined by the competition between near-cognate and cognate aminoacyl-tRNAs,^[21] a high modification level increases the affinity of the correct tRNA to the ribosome, which may allow faster discrimination.^[22] This reduces the ribosome step time, which in turn may increase protein synthesis rates.

In summary, we have reported the parallel quantification of 12 modified nucleosides in tRNA ensembles from various porcine and murine tissues and showed that the overall modification content varies substantially. Furthermore, we provide evidence that the modification level correlates with the in vitro protein synthesis capacity, suggesting that the extent to which the tRNA ensemble is chemically modified modulates the translational efficiency. Our data show that the tRNA modification level is another layer of information that programs cells in terms of their translational potency.

Received: May 16, 2012

Revised: July 26, 2012

Published online: October 4, 2012

Keywords: isotopic labeling · mass spectrometry · RNA modification · translation · tRNA

- [1] M. Ibba, D. Söll, *Science* **1999**, 286, 1893–1897.
- [2] a) N. Fischer, A. L. Konevega, W. Wintermeyer, M. V. Rodnina, H. Stark, *Nature* **2010**, 466, 329–333; b) T. M. Schmeing, V. Ramakrishnan, *Nature* **2009**, 461, 1234–1242.
- [3] a) H. Grosjean, *Fine-Tuning of RNA Functions by Modification and Editing*, Springer, Heidelberg, **2005**; b) F. Jühling, M. Mörl,

- K. Hartmann Roland, M. Sprinzl, F. Stadler Peter, J. Pütz, *Nucleic Acids Res.* **2009**, 37, D159–D162.
- [4] Y. Motorin, M. Helm, *Biochemistry* **2010**, 49, 4934–4944.
- [5] a) P. F. Agris, F. A. Vendeix, W. D. Graham, *J. Mol. Biol.* **2007**, 366, 1–13; b) A. Ambrogelly, S. Palioura, D. Söll, *Nat. Chem. Biol.* **2007**, 3, 29–35.
- [6] a) L. B. Jenner, N. Demeshkina, G. Yusupova, M. Yusupov, *Nat. Struct. Mol. Biol.* **2010**, 17, 555–560; b) A. L. Konevega, N. G. Soboleva, V. I. Makhno, A. V. Peshekhonov, V. I. Katunin, *Mol. Biol.* **2006**, 40, 597–610; c) K. Nakanishi, L. Bonnefond, S. Kimura, T. Suzuki, R. Ishitani, O. Nureki, *Nature* **2009**, 461, 1144–1148; d) L. A. Sylvers, K. C. Rogers, M. Shimizu, E. Ohtsuka, D. Söll, *Biochemistry* **1993**, 32, 3836–3841.
- [7] J. F. Atkins, G. R. Björk, *Microbiol. Mol. Biol. Rev.* **2009**, 73, 178–210.
- [8] a) T. Brückl, D. Globisch, M. Wagner, M. Müller, T. Carell, *Angew. Chem.* **2009**, 121, 8074–8077; *Angew. Chem. Int. Ed.* **2009**, 48, 7932–7934; b) J. J. Dalluge, T. Hashizume, J. A. McCloskey, *Nucleic Acids Res.* **1996**, 24, 3242–3245.
- [9] M. Yarus, *Science* **1982**, 218, 646–652.
- [10] a) E. M. Phizicky, J. D. Alfonzo, *FEBS Lett.* **2010**, 584, 265–271; b) M. Helm, *Nucleic Acids Res.* **2006**, 34, 721–733.
- [11] a) J. B. Plotkin, H. Robins, A. J. Levine, *Proc. Natl. Acad. Sci. USA* **2004**, 101, 12588–12591; b) K. A. Dittmar, J. M. Goodenbour, T. Pan, *PLoS Genet.* **2006**, 2, 2107–2115.
- [12] J. C. Waterlow, *Exp. Physiol.* **1984**, 69, 409–438.
- [13] D. Globisch, D. Pearson, A. Hienzs, T. Brückl, M. Wagner, I. Thoma, P. Thumbs, V. Reiter, A. C. Kneuttinger, M. Müller, S. A. Sieber, T. Carell, *Angew. Chem.* **2011**, 123, 9913–9916; *Angew. Chem. Int. Ed.* **2011**, 50, 9739–9742.
- [14] a) Y. Motorin, M. Helm, *Wiley Interdiscip. Rev. RNA* **2011**, 2, 611–631; b) J. A. Kowalak, J. J. Dalluge, J. A. McCloskey, K. O. Stetter, *Biochemistry* **1994**, 33, 7869–7876.
- [15] a) T. J. Siard, J. R. Katze, W. R. Farkas, *Neurochem. Res.* **1989**, 14, 1159–1164; b) L. Szabo, S. Nishimura, W. R. Farkas, *Biofactors* **1988**, 1, 241–244.
- [16] C. M. Castleberry, P. A. Limbach, *Nucleic Acids Res.* **2010**, 38, e162.
- [17] a) A. Suryawan, P. M. J. O'Connor, J. A. Bush, H. V. Nguyen, T. A. Davis, *Amino Acids* **2009**, 37, 97–104; b) D. F. Goldspink, S. E. Lewis, F. J. Kelly, *Biochem. J.* **1984**, 217, 527–534; c) M. Winick, A. Noble, *Dev. Biol.* **1965**, 12, 451–466; d) J. W. Frank, J. Escobar, A. Suryawan, S. R. Kimball, H. V. Nguyen, L. S. Jefferson, T. A. Davis, *J. Nutr.* **2005**, 135, 1374–1381.
- [18] R. J. Jackson, S. Napthine, I. Brierley, *RNA* **2001**, 7, 765–773.
- [19] V. Reiter, D. M. S. Matschkal, M. Wagner, D. Globisch, A. C. Kneuttinger, M. Müller, T. Carell, *Nucleic Acids Res.* **2012**, 40, 6235–6240.
- [20] C. Frezza, S. Cipolat, L. Scorrano, *Nat. Protoc.* **2007**, 2, 287–295.
- [21] J. Ling, N. Reynolds, M. Ibba, *Annu. Rev. Microbiol.* **2009**, 63, 61–78.
- [22] a) A. L. Konevega, N. G. Soboleva, V. I. Makhno, Y. P. Semeni-
kov, W. Wintermeyer, M. V. Rodnina, V. I. Katunin, *RNA* **2004**,
10, 90–101; b) O. Allnér, L. Nilsson, *RNA* **2011**, 17, 2177–2188;
c) V. Dao, R. Guenther, A. Malkiewicz, B. Nawrot, E. Sochacka,
A. Kraszewski, J. Jankowska, K. Everett, P. F. Agris, *Proc. Natl.
Acad. Sci. USA* **1994**, 91, 2125–2129; d) J. Urbonavičius, J. M. B.
Durand, G. R. Björk, *J. Bacteriol.* **2002**, 184, 5348–5357.

4.3 „Statistische Analyse der tRNA-Modifikationssets von Krebszelllinien als Basis für eine Klassifizierung“

In diesem Projekt sollten die tRNA-Modifikationssets von humanen Krebszelllinien mit der von Daniel Globisch und Tobias Brückl entwickelten Methode zur parallelen Quantifizierung von RNA-Modifikationen untersucht werden. Ziel war es, eine Methode zu entwickeln, mit der über eine Analyse der tRNA-Modifikationssets eine Klassifikation unbekannter Krebszelllinien möglich würde.

Ich konnte hier zeigen, dass jede Krebszelllinie ein für diese Zelllinie charakteristisches tRNA-Modifikationsset besitzt, welches ähnlich wie ein Fingerabdruck prinzipiell für eine Identifikation der Zelllinie genutzt werden kann. Wie die von mir durchgeführte statistische Analyse ergab, weist das tRNA-Set einer Krebszelllinie daneben im Vergleich zu primären Zelllinien eine größere Varianz der einzelnen Modifikationen auf. Dies erlaubt eine Unterscheidung von Krebszelllinien und nicht-malignen Zelllinien allein anhand ihres tRNA-Modifikationssets. Eine Aussage über das Herkunftsorgan einer Krebszelllinie ist dagegen anhand der tRNA-Modifikationen nicht möglich.

Alle hier beschriebenen Experimente und Analysen sind von mir eigenständig durchgeführt worden.

Die Ergebnisse dieser Studie sind bisher nicht publiziert worden.

Statistische Analyse der tRNA-Modifikationssets von Krebszelllinien als Basis für eine Klassifizierung

Während einzelne tRNA-Modifikationen und ihre biologische Funktion z.T. genau untersucht und verstanden sind,^[2, 428-432] wurde das systemische Verhalten der modifizierten Nukleoside einer tRNA oder eines tRNA-Sets als Gesamtheit bisher wenig erforscht.^[97-98, 433-443] Dabei richtet sich das Interesse in der modernen Biochemie immer mehr auf das Verständnis von biologischen Systemen als Ganzen.^[444-445] So wird beispielsweise in der Proteomik die Gesamtheit aller in einer Zelle unter definierten Bedingungen und zu einem definierten Zeitpunkt vorliegenden Proteine, parallel analysiert.^[446] Dies ermöglicht fundamental neue Einblicke in das Wechselspiel der Proteine untereinander und ein tieferes Verständnis des Systems als Ganzem.^[445, 447-449]

Frühe Versuche, diesen Ansatz auf die Erforschung der tRNA-Modifikationen zu übertragen, waren oft nicht genau genug. Meistens wurde in diesen Studien versucht, die modifizierten Nukleoside chromatographisch aufzutrennen und dann anhand ihrer UV-Absorption parallel zu quantifizieren.^[450-451] Die Empfindlichkeit dieser Methode ist nicht besonders groß, da eine vollständige Auftrennung der Modifikationen kaum möglich ist und viele modifizierte Nukleoside außerdem nur in sehr geringer Menge vorkommen.^[451] Damit überlappen sich ihre UV-Signale oft mit denen der kanonischen Nukleoside, bzw. sie sind erst gar nicht zu detektieren.

4.3.1 Vorarbeiten

Im Arbeitskreis von Prof. Carell ist deshalb von Daniel Globisch und Tobias Brückl eine neue und verbesserte Methode zur parallelen Quantifizierung von modifizierten RNA-Nukleosiden entwickelt und publiziert worden.^[97] Die Modifikationen werden bei dieser Methode massenspektrometrisch quantifiziert; das Verfahren ist damit im Vergleich zur UV-basierten Quantifizierung um ein Vielfaches sensitiver.^[452-454] Die von Carell und Mitarbeitern entwickelte Methode^[97] erlaubt als Isotopenverdünnungsmethode außerdem eine exakte Quantifizierung der absoluten Nukleosidmengen in der Probe.^[436, 440] Dafür werden mit stabilen Isotopen markierte Nukleosid-Standards, welche zuvor von Carell und Mitarbeitern synthetisiert wurden,^[4, 98, 435, 455] zur Probe hinzugegeben. Die Nukleosid-Standards besitzen dasselbe chromatographische Verhalten und dieselben Ionisationseigenschaften wie die natürlichen Nukleoside, haben jedoch ein anderes Molekulargewicht. Damit wird eine separate Integration der Massensignale der natürlichen Nukleoside und der isotopenmarkierten Standards möglich. Durch Vergleich der Integrale lassen sich dann die natürlichen Modifikationen in der Probe mit Hilfe einer Eichgeraden genau quantifizieren.

Die Leistungsfähigkeit dieser neuen Quantifizierungsmethode wurde zunächst durch die Analyse von tRNA-Modifikationen in humanen Krebszelllinien, *E.coli* und Schweineleber demonstriert.^[97] Dabei wurden die genauen Level von sechs Adenosinmodifikationen (siehe Abbildung 19) bestimmt.

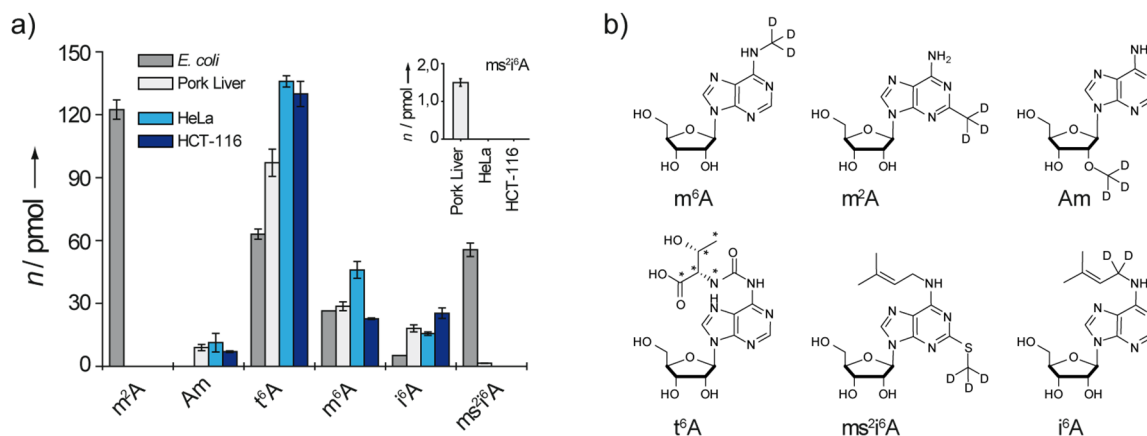


Abbildung 19. Massenspektrometrische Quantifizierung von tRNA-Modifikationen in *E. coli*, Schweineleber und den humanen Krebszelllinien HeLa und HCT-116 nach der von Carell und Mitarbeitern entwickelten Isotopenverdünnungsmethode.^[97] a) Stoffmengen der modifizierten Nukleoside in pmol; die Fehlerbalken ergeben sich aus den Standardabweichungen von mindestens drei biologischen Replikaten. b) Strukturformeln der quantifizierten Nukleoside.

Es zeigte sich, dass sich Vorkommen und Häufigkeit dieser Modifikationen bei den in dieser Studie untersuchten Organismen stark unterscheiden. Daneben wurde bei einzelnen Modifikationen auch zwischen den humanen Krebszelllinien HeLa und HCT-116 ein erheblicher Unterschied beobachtet.^[97] Ob ein solcher Unterschied bei vielen der in der tRNA vorkommenden Modifikationen existiert und ob sich diese Unterschiede für eine Klassifikation von Krebszelllinien nutzen lassen, war zu diesem Zeitpunkt noch weitgehend unbekannt.^[450, 456-457]

4.3.2 Zur Klassifizierung von Krebszelllinien

Die einfachste Art, Krebszelllinien zu klassifizieren, ist die Gruppierung der Zelllinien nach Herkunftsorgan oder –gewebe. Weitaus genauer ist eine Klassifikation anhand des Genexpressions-^[458-469] oder Enzymaktivitätsprofils.^[470-472] Oft folgt aus der Analyse von Genexpressions- oder Enzymaktivitätsdaten nicht nur die „klassische“ Gruppierung der Krebszelllinien nach den Herkunftsorganen, sondern die Methoden erlauben darüber hinaus die Identifikation von Untergruppen innerhalb der durch die Herkunftsorgane oder –gewebe gebildeten Hauptgruppen, welche auf andere Weise nicht so offensichtlich zu erkennen sind.^[458, 462, 471] Diese über ihre

biochemischen Eigenschaften identifizierten Untergruppen besitzen im besten Fall auch gemeinsame histologische oder zelluläre Eigenschaften, wie z.B. eine ähnliche Malignität oder Invasivität.^[471]

Für die Identifizierung von Zelllinien, die sich aufgrund ihrer starken Ähnlichkeit zu einer Gruppe zusammenfassen lassen, ist ein professionelles *Data-Mining* in den Genexpressions- oder Enzymaktivitätsdaten notwendig. Darunter versteht man die statistische Analyse eines großen Datenbestandes mit dem Ziel, neue Muster zu erkennen. Die sogenannte Clusteranalyse ist ein häufig angewendetes statistisches Analyseverfahren, um ähnliche Strukturen oder Objekte in einem Datenbestand zu identifizieren.^[473] Die nach diesem Verfahren gefundenen Gruppen von ähnlichen Objekten werden als *Cluster* bezeichnet, die Zuordnung von Objekten zu diesen Clustern bezeichnet man als *Clustering*. Mit Hilfe der Clusteranalyse ist es möglich, Krebszelllinien anhand ihres Genexpressions-^[458-463] oder Enzymaktivitätsprofils ^[470-472] genauer zu klassifizieren.

Das außerordentliche Interesse an einer solchen biochemischen Klassifikation von Krebsarten bzw. Krebszelllinien beruht auf der Hoffnung, dass sich ausgehend von diesen Informationen - Stichwort *personalisierte Medizin* - die Heilungschancen der betroffenen Patienten deutlich erhöhen lassen.^[474]

So sollten Krebszelllinien, deren Genexpressions- oder Enzymaktivitätsprofil sich stark ähnelt, im besten Fall auch auf Krebsmedikamente ähnlich reagieren.^[459] Kennt man daher die Wirkung eines Medikaments auf eine Zelllinie eines Clusters, so könnte man daraus die Wirkung auf die anderen Zelllinien desselben Clusters ableiten. Ebenso könnten die bei Biopsien gewonnenen Proben weitaus genauer als bisher klassifiziert werden, so dass eine zielgerichtete Behandlung des Patienten möglich wird.

4.3.3 Quantifizierung von tRNA-Modifikationen humaner Krebszelllinien

Das Wissen um die Möglichkeiten der statistischen Datenanalyse und ihre Bedeutung für die Klassifikation von Krebszelllinien, zusammen mit der Beobachtung von Unterschieden in den Modifikationssets der Zelllinien HeLa und HCT-116 ^[97] führten zu der Idee, eine Methode zu entwickeln, mit der über eine Analyse des tRNA-Modifikationssets eine Klassifikation unbekannter Krebszelllinien möglich wäre.

In einem ersten Schritt sollten dazu die Modifikationslevel von verschiedenen Krebszelllinien mit der im Arbeitskreis entwickelten Methode zur parallelen Quantifizierung von modifizierten RNA-Nukleosiden bestimmt werden.^[97] Durch eine Clusteranalyse ^[473] dieser Modifikationslevel sollten dann Gruppen bzw. Cluster ähnlicher Zelllinien identifiziert werden. Die Qualität des Clusterings sollte abschließend anhand bereits publizierter Genexpressions-^[458-463] oder Enzymaktivitätsdaten ^[470-472]

überprüft werden. Im Idealfall sollte mit dieser Methode zumindest eine Identifikation des Herkunftsorgans unbekannter Krebszelllinien möglich sein.

4.3.3.1 Auswahl der Zelllinien

Bei der Auswahl der Krebszelllinien wurde darauf geachtet, dass diese zuvor schon hinreichend charakterisiert worden sind. Dies sollte eine Validierung der gefundenen Cluster bei der geplanten Clusteranalyse erleichtern und damit eine Abschätzung der Methodenqualität erlauben. Besonders gut untersucht und klassifiziert sind die sogenannten NCI60-Zelllinien.^[475-482] Es handelt sich dabei um 60 Krebszelllinien, welche vom *National Cancer Institute* (NCI) für ein Wirkstoff-Screening mit 70.000 potentiellen Antikrebswirkstoffen verwendet wurden und die daher pharmakologisch genau charakterisiert sind.^[482] Weiterhin wurden von diesen Zelllinien auch die Genexpressionsprofile analysiert und mit dem jeweiligen Phänotyp korreliert.^[459-460] Für eine Auswahl der NCI60- Zelllinien liegen außerdem Daten zu den Enzymaktivitätsprofilen vor, die von Cravatt und Mitarbeitern in mehreren Studien veröffentlicht worden sind.^[471-472, 483] Ich habe deshalb für die hier beschriebene Studie Zelllinien bevorzugt, die zu den NCI60-Zelllinien gehören und deren Enzymaktivitätsprofil möglichst auch untersucht worden ist (Tabelle 1).

Brust	Gewebetyp	Krebstyp	NCI60-Zelllinie	ABPP-Daten
MDA-MB-231	Brustdrüse	Adenokarzinom	+	+
MCF-7	Brustdrüse	Adenokarzinom	+	+
T-47D	Brustdrüse	Duktales Karzinom	+	+
BT-549	Brustdrüse	Duktales Karzinom	+	--
Haut				
SK-MEL-2	Melanocyten	Malignes Melanom	+	+
SK-MEL-5	Melanocyten	Malignes Melanom	+	--
SK-MEL-28	Melanocyten	Malignes Melanom	+	--
IGR-1	Melanocyten	Malignes Melanom	--	--
Leber				
SK-HEP-1	Leber	Adenokarzinom	--	--
Primäre Zelllinien				
HMEC	Brust, Epithelzellen	--	--	--
HEMa	Haut, Melanocyten	--	--	--

Tabelle 1. Liste der in dieser Studie untersuchten Zelllinien. ABPP: *Activity-Based Protein Profiling*, von Cravatt und Mitarbeitern entwickelte Methode zur Bestimmung von Enzymaktivitätsprofilen.^[470]

Für die Analyse der tRNA-Modifikationen wurden von mir insgesamt neun humane Krebszelllinien ausgewählt. Darunter befinden sich vier Brustkrebszelllinien, vier Hautkrebszelllinien und eine

Leberkrebszelllinie. Sieben dieser Zelllinien sind Teil der NCI60-Zelllinien, und von vier der sieben Zelllinien ist auch das Enzymaktivitätsprofil ^[471-472] untersucht worden. Zum Vergleich wurden daneben zwei primäre humane Zelllinien untersucht. Eine dieser primären Zelllinien stammt aus Epithelzellen der Brust (HMEC), die andere aus Melanocyten der Haut (HEMa) (siehe Tabelle 1).

4.3.3.2 Auswahl der tRNA-Modifikationen

Eine Klassifikation von Krebszelllinien anhand ihrer tRNA-Modifikationssets sollte umso genauer sein, je mehr Modifikationen in die Analyse miteinbezogen werden. Die im Arbeitskreis Carell entwickelte Methode zur Quantifizierung von modifizierten RNA-Nukleosiden erlaubt im Prinzip die parallele Quantifizierung aller tRNA-Modifikationen.^[97] In der Praxis ist es jedoch sinnvoll, sich auf eine repräsentative Auswahl dieser Modifikationen zu beschränken, da jede der nach dieser Methode untersuchte Modifikation zumindest als stabiles Isotopolog synthetisiert werden muss. Wir haben uns daher auf die Quantifizierung von zehn repräsentativen Modifikationen beschränkt (siehe Abbildung 20), von denen m¹G und Gm aus technischen Gründen zusammen quantifiziert wurden.

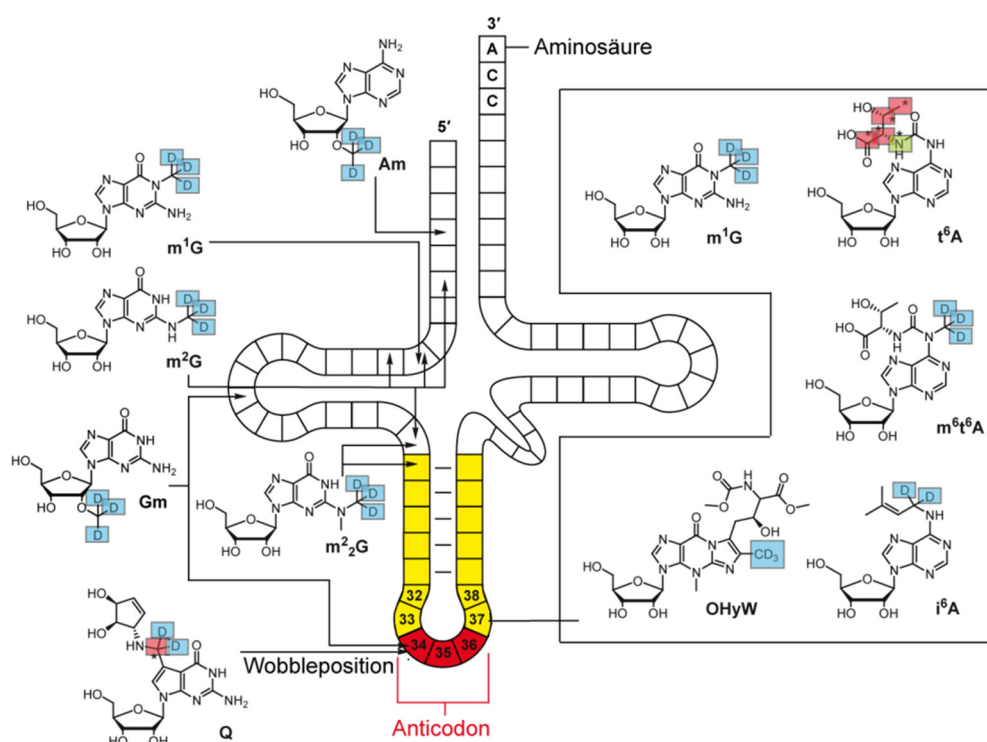


Abbildung 20. Kleeblattförmige Sekundärstruktur der tRNA (rot: Anticodon, gelb: Anticodonarm) mit den Strukturformeln und Positionen der in dieser Studie quantifizierten Modifikationen. Blau markierte Atome: Deuterium; rot markierte Atome: ¹³C; grün markiertes Atom: ¹⁵N.

Diese zehn Modifikationen wurden in ihrer natürlichen Form und auch als isotopenmarkierte Referenznukleoside von Tobias Brückl, Daniel Globisch, Antje Hienzsch und Florian Klepper in der Arbeitsgruppe Carell synthetisiert.^[4, 98, 435, 455] Die Mehrheit der hier quantifizierten Modifikationen findet man im erweiterten Anticodon der tRNA und dort besonders in der konservierten Purinposition 37 direkt neben dem Anticodon.^[7, 484-489] Diese Modifikationen sorgen im Wesentlichen für die hohe Genauigkeit der Translation.^[11, 490] Die restlichen Modifikationen findet man an verschiedenen anderen Positionen der tRNA-Struktur.^[7, 484-489] Dort sind sie hauptsächlich an der Stabilisierung der tRNA-Tertiärstruktur beteiligt^[13, 491-493] oder verbessern die substratspezifische Bindung zur entsprechenden tRNA-Aminoacylsynthetase.^[494-500]

4.3.3.3 Extraktion der tRNA und Quantifizierung der Modifikationen

Zu Beginn des Projektes wurden die oben genannten humanen Krebszelllinien von mir in Kultur genommen. Durch mehrmaliges Splitting war es leicht möglich, genügend Zellen für die tRNA-Extraktion zu erhalten. Diese erfolgte nach einer von mir für Zellkulturproben adaptierten Methode, welche ursprünglich von Tobias Brückl in der Arbeitsgruppe Carell für die tRNA-Extraktion aus Schweinegeweben entwickelt worden war.^[97, 440]

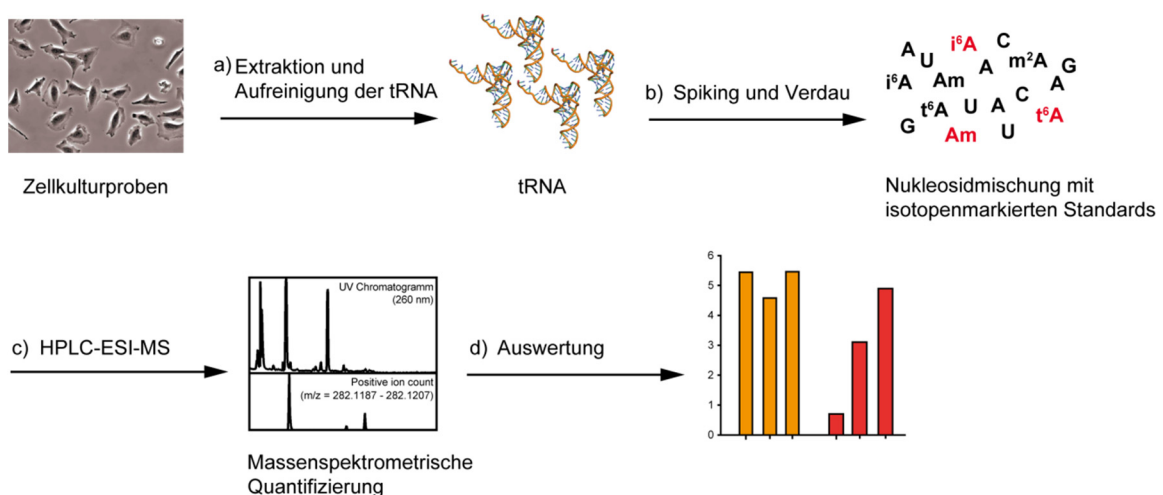


Abbildung 21. Schematische Darstellung der von Carell und Mitarbeitern entwickelten Methode zur exakten massenspektrometrischen Quantifizierung von modifizierten RNA-Nukleosiden nach der Isotopenverdünnungsmethode.^[97, 440]

Die Extraktion beginnt mit dem Aufschluss der Zellen durch eine wässrige Phenollösung, gefolgt von einer Phenol-Chloroform-Extraktion zur Abtrennung der Proteine. Bei den nachfolgenden

Fällungsschritten wird insbesondere die genomische DNA abgetrennt. Nach der abschließenden Ionenaustauschchromatographie konnte die tRNA in hoher Reinheit isoliert werden.

Um die tRNA-Modifikationen massenspektrometrisch quantifizieren zu können, war es erforderlich, die tRNA enzymatisch zu den Nukleosiden zu verdauen. Dann wurden die stabil isotoopenmarkierten Referenznukleoside in genau bekannter Menge zu den Proben zugegeben. Diese wurden im Anschluss über ein *UltiMate 3000* HPLC-System in Kombination mit einem *Orbitrap XL* Massenspektrometer analysiert.^[97, 440] Eine schematische Darstellung des gesamten hier beschriebenen Ablaufs findet sich in Abbildung 21.

Von jeder Zelllinie wurden zwei biologische Replikate analysiert. Bei diesen handelte es sich um Zellen, die an unterschiedlichen Tagen und von unterschiedlichen Passagen geerntet worden waren. Es wurden je Probe mindestens drei separate HPLC-ESI-MS-Messungen durchgeführt (technische Replikate), wobei jeweils 12 µg tRNA verdaut und analysiert wurden.

4.3.3.4 Ergebnisse der Quantifizierung und Diskussion

Die Ergebnisse der HPLC-ESI-MS-Messungen finden sich als Übersicht in Abbildung 22. Zum besseren Verständnis sind die Messergebnisse dort in *Modifikationen pro 1000 tRNAs* angegeben. (Die Anzahl einer bestimmten Modifikation pro 1000 tRNAs kann leicht aus der in 12 µg tRNA gefundenen Stoffmenge dieser Modifikation berechnet werden. Dazu wird eine durchschnittliche molare Masse der tRNA-Moleküle von $M = 25.000 \text{ g mol}^{-1}$ angenommen). Die gewählte Farbkodierung dient der Übersichtlichkeit; rot markierte Felder zeigen eine überdurchschnittlich hohe Modifikationsdichte einer bestimmten Modifikation an. Gelb bzw. grün markierte Felder stehen für eine mittlere bzw. unterdurchschnittliche Modifikationsdichte. Da sich m¹G und Gm weder in der Retentionszeit noch der Masse unterscheiden, wurden sie zusammen quantifiziert.

Betrachtet man das Resultat der Messungen, so findet man bei allen untersuchten Modifikationen z.T. erhebliche Unterschiede zwischen den Zelllinien. Wie wir es auch nach unseren ersten Messungen vermutet haben,^[97] besitzt jede Zelllinie tatsächlich ein individuelles Set an tRNA-Modifikationen. Dieses unterscheidet sich deutlich vom Modifikationsset anderer Zelllinien. Allerdings ist auf den ersten Blick und ohne statistische Analyse keine modifikationsbasierte Klassifizierung der Krebszelllinien möglich. Es besteht also offenbar kein offensichtlicherer Unterschied zwischen dem tRNA-Modifikationsset einer Brustkrebs- und einer Hautkrebszelllinie, als z.B. zwischen den Modifikationssets zweier Brustkrebszelllinien.

Vergleicht man die tRNA-Modifikationssets von Krebs- und primären Zelllinien, so fällt besonders der geringe tRNA-Modifikationsgrad der primären Zelllinie HEMa auf. Der Modifikationsgrad ist hier nicht nur niedriger als bei den Krebszelllinien, er ist auch niedrig im Vergleich zur zweiten primären Zelllinie HMEC. Der durchschnittliche HMEC-Modifikationsgrad liegt dagegen, verglichen mit den Krebszelllinien, im unteren Mittelfeld.

%o tRNA	t6A	m6t6A	i6A	OH-yW	Q	m1G+Gm	m22G	m2G	Am	
MDA-MB-231	272	18	37	10	62	898	579	992	14	Brustkrebs
MCF-7	278	19	50	28	65	879	608	1150	15	
T-47D	247	23	40	25	78	941	526	1154	26	
BT-549	247	23	35	tr	105	862	538	938	15	
SK-MEL-2	242	15	39	18	74	755	540	848	26	Hautkrebs
SK-MEL-5	260	17	52	24	33	844	568	1003	21	
SK-MEL-28	264	19	46	12	51	870	586	1026	25	
IGR-1	276	21	45	12	88	839	549	909	21	
SK-HEP-1	228	22	38	25	54	843	556	901	26	Leberkrebs
HMEC	257	19	41	19	0	794	554	929	17	Primäre Zelllinien
HEMa	153	12	29	9	12	450	326	689	11	

Abbildung 22. Ergebnisse der HPLC-MS-basierten Quantifizierung von 10 modifizierten Nukleosiden in der tRNA von 9 Krebszelllinien und zwei primären Zelllinien. Aus technischen Gründen wurden die Guanosinderivate m¹G und Gm zusammen quantifiziert. Die jeweils gemessene Stoffmenge wurde zum besseren Verständnis in *Modifikationen pro 1000 tRNAs* umgerechnet. Der für die Spalten, d.h. für die Modifikationsdichte pro Zelllinie gewählte Farbcode reicht von dunkelgrün (sehr niedrig) bis dunkelrot (sehr hoch).

Es ist weiterhin auffällig, dass der individuelle Modifikationsgrad der verschiedenen Modifikationen bei den Krebszelllinien offenbar stärker als bei den primären Zelllinien schwankt. So besitzt beispielsweise die Hautkrebszelllinie IGR-1 den zweithöchsten Modifikationsgrad bezüglich der Modifikation t⁶A, die Modifikationsdichte von Hydroxywybutosin (HO-yW) ist im Vergleich zu den andern Zelllinien jedoch besonders niedrig.

Im Gegensatz dazu ist bei den primären Zelllinien HEMa und HMEC der individuelle Modifikationsgrad bei allen untersuchten Modifikationen im Vergleich zu den anderen Zelllinien entweder besonders niedrig (HEMa) oder leicht unterdurchschnittlich (HMEC). Wie wir bereits zeigen konnten, schwankt der durchschnittliche Modifikationsgrad eines tRNA-Sets bei Säugetieren zwischen den verschiedenen Organen bzw. Geweben relativ stark.^[440] Dagegen weist das tRNA-Set eines bestimmten Organs im Vergleich zu den tRNA-Sets anderer Organe jeweils eine ähnliche relative Modifikationsdichte bezüglich der verschiedenen Modifikationen auf.^[440] So findet man z.B. in der Leber bei allen untersuchten Modifikationen eine besonders hohe relative Modifikationsdichte. Auch in den hier

untersuchten tRNA-Sets der primären Zelllinien findet sich bei den einzelnen Modifikationen jeweils ein im Vergleich zu anderen Zelllinien ähnlicher relativer Modifikationsgrad. Die innerhalb der tRNA-Sets von Krebszelllinien beobachteten starken Schwankungen der individuellen Modifikationsgrade deuten deshalb möglicherweise auf gestörte biosynthetische Pfade hin, die zu einer krankhaft verminderten oder verstärkten Biosynthese einzelner Modifikationen führen. Eine vergleichbare Störung der biosynthetischen Pfade der Zelle führt auch zu dem für Krebszellen typischen anomalen Energiestoffwechsel.^[501-503]

Die Messwerte aus Abbildung 22 sind zum besseren Verständnis nochmals in Abbildung 23 als Balkendiagramme dargestellt. Wir sehen wieder, dass alle in dieser Studie untersuchten methylierten Guanosinderivate, aber auch t⁶A zu den häufig in tRNAs gefundenen Modifikationen gehören.

Die Modifikationsdichte liegt hier zwischen durchschnittlich 958 Modifikationen pro 1000 tRNAs bei m²G und 248 Modifikationen pro 1000 tRNAs bei t⁶A. Die Modifikationsdichte der im Anticodon vorkommenden Modifikationen ist generell geringer, sie liegt dort zwischen durchschnittlich 57 Modifikationen pro 1000 tRNAs bei Queuosin (Q) und nur 18 Modifikationen pro 1000 tRNAs bei HO-yW. Wie man leicht sieht, sind die Schwankungen der Modifikationsdichte, die man bei Am, HO-yW und Q beobachtet, besonders groß. Lässt man die Zelllinie HEMa wegen ihres außergewöhnlich niedrigen Modifikationsgrads außen vor, so beträgt der prozentuale Unterschied zwischen kleinster und größter Modifikationsdichte hier 90% (Am), 190% (HO-yW) bzw. 220% (Q). Bei den Modifikationen m⁶t⁶A und i⁶A beträgt dieser Unterschied immer noch jeweils 50%. Es fällt auf, dass diese Modifikationen abgesehen von Am alle im erweiterten Anticodon der tRNA vorkommen und dort besonders für die hohe Genauigkeit der Translation sorgen (siehe auch Abbildung 20).^[11, 490] Bei den hauptsächlich an der Stabilisierung der tRNA-Tertiärstruktur beteiligten Modifikationen m²₂G und m²G beobachtet man dagegen nur noch einen prozentualen Unterschied von 20% bzw. 40% zwischen der geringsten und höchsten gemessenen Modifikationsdichte.^[504]

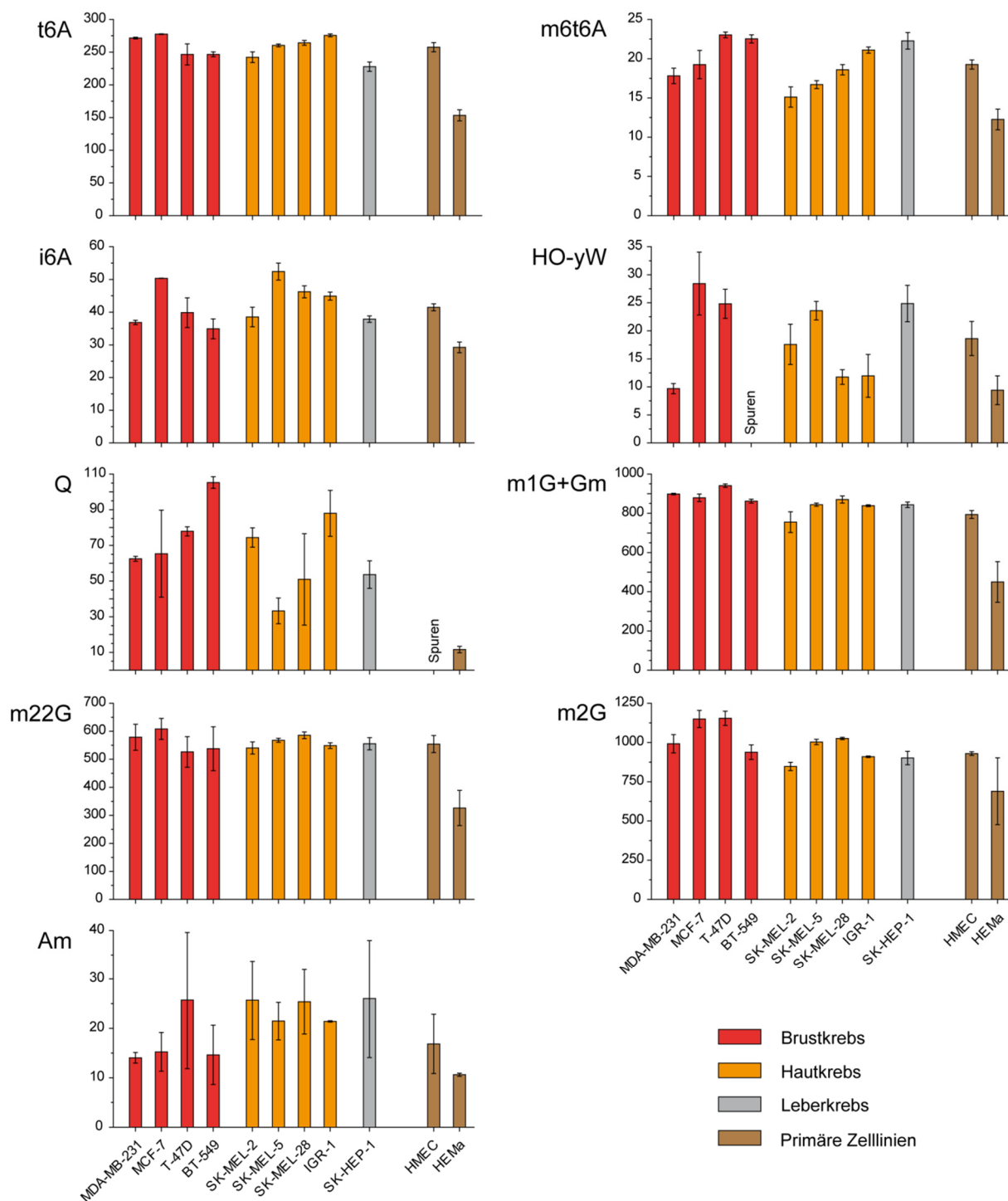


Abbildung 23. Ergebnisse der HPLC-MS-basierten Quantifizierung von 10 modifizierten Nukleosiden in der tRNA von 9 Krebszelllinien und zwei primären Zelllinien; Darstellung als Balkendiagramme (vgl. Abbildung 22). Aus technischen Gründen wurden die Guanosinderivate m¹G und Gm zusammen quantifiziert. Alle Werte sind in *Modifikationen pro 1000 tRNAs* angegeben; die Fehlerbalken ergeben sich aus der Standardabweichung von zwei bis drei biologischen Replikaten. Rot: Brustkrebszelllinien; orange: Hautkrebszelllinien; grau: Leberkrebszelllinie; braun: Primäre Zelllinien.

4.3.4 Statistische Analyse der tRNA-Modifikationssets

Im Folgenden wird beschrieben, wie über eine statistische Analyse Ähnlichkeiten zwischen den Modifikationssets der verschiedenen Zelllinien identifiziert wurden. Dann wird gezeigt, ob und wie sich diese Ähnlichkeiten nutzen lassen, um mit Hilfe einer hierarchischen Clusteranalyse^[505] (Krebs)zelllinien zu klassifizieren.

4.3.4.1 Die Pearson-Korrelation und der Pearson-Korrelationskoeffizient als Ähnlichkeitsmaß zwischen Zahlen- bzw. Datenreihen

Die Ähnlichkeit zweier Modifikationssets wurde von mir über die Pearson-Korrelation berechnet.^[506] Der dabei erhaltene Pearson-Korrelationskoeffizient beschreibt das Ausmaß der Ähnlichkeit zwischen den betrachteten Modifikationssets. Der Pearson-Korrelationskoeffizient ist ein häufig verwendetes statistisches Hilfsmittel, um das Ausmaß der Ähnlichkeit zweier Zahlenreihen (bzw. Vektoren) $x = \{x_1, x_2, \dots, x_n\}$ und $y = \{y_1, y_2, \dots, y_n\}$ zu beschreiben.^[507-510] Er ist folgendermaßen definiert:^[506]

$$r = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i - x_m}{\sigma_x} \right) \left(\frac{y_i - y_m}{\sigma_y} \right)$$

Dabei ist x_m der Mittelwert der Werte von x und σ_x die Standardabweichung dieser Werte. Der Pearson-Korrelationskoeffizient liegt damit immer zwischen -1 und 1, wobei $r = 1$ eine perfekte Übereinstimmung und $r = 0$ das Fehlen jeglicher Korrelation zwischen den Zahlenreihen bedeutet. Ein Pearson-Korrelationskoeffizient von $r = -1$ beschreibt zwei Zahlenreihen, die perfekt antisymmetrisch sind.

Es gibt mehrere Möglichkeiten, sich die durch den Pearson-Korrelationskoeffizienten beschriebene Ähnlichkeit zu veranschaulichen: Trägt man beide Zahlenreihen in einem Koordinatensystem gegeneinander auf, so nähert sich r umso mehr der 1, je besser man eine Gerade durch die Punkte (x_1, y_1) , (x_2, y_2) etc. legen kann. Man kann sich jede Zahlenreihe auch als eine Kurve in einem Koordinatensystem mit den Punkten $(1, x_1)$, $(2, x_2)$ etc. vorstellen. Dann ist die Pearson-Korrelation zwischen zwei Zahlenreihen umso höher, je ähnlicher die Form ihrer Kurven ist.^[511]

Es ist wichtig zu wissen, dass der Pearson-Korrelationskoeffizient in der hier verwendeten Form invariant gegenüber Lineartransformationen der Zahlen- bzw. Datenreihen ist. Vergleicht man daher zwei Zahlenreihen, so erhält man denselben Wert für r auch dann, wenn man bei einer (oder bei beiden) Zahlenreihe(n) jeden Wert x_i mit derselben Zahl multipliziert oder zu jedem Wert dieselbe

Zahl addieren (oder subtrahieren) würde. Stellen wir uns wieder durch die Zahlenreihen beschriebene Kurven in einem Koordinatensystem vor. Dann wird nur die Form der Kurven miteinander verglichen, nicht jedoch ihre absolute Position im Koordinatensystem. Dies ist in unserem Fall besonders wichtig für die korrekte Einordnung der primären HEMA-Zelllinie. Wie oben gesehen, weist die tRNA von HEMA bei allen betrachteten Modifikationen im Vergleich zu den anderen Zelllinien einen besonders niedrigen Modifikationsgrad auf. Es kann nicht mit letzter Sicherheit ausgeschlossen werden, dass dieses außergewöhnlich niedrige Modifikationsniveau durch die Analyse einer zu geringen tRNA-Probenmenge zustande kam. Dennoch würde dies bei der Clusteranalyse mit Hilfe der Pearson-Korrelation keine Rolle spielen. Wie oben ausgeführt, hängt der Pearson-Korrelationskoeffizient zweier Modifikationssets nicht von den absoluten Mengen der Modifikationen ab. Er wird stattdessen über das Verhältnis der einzelnen Modifikationen einer Zelllinie zueinander bestimmt.

4.3.4.2 Normierung der Modifikationslevel

Um mit den bei unseren Messungen erhaltenen Daten eine Clusteranalyse durchführen zu können, war zunächst eine Normalisierung der für die verschiedenen Modifikationen erhaltenen Messwerte notwendig. Dies ist leicht verständlich, wenn man bedenkt, dass der Pearson-Korrelationskoeffizient der Zahlenreihen $\{10; 300; 40; 1500; 90\}$ und $\{20; 330; 50; 1600; 80\}$ annähernd eins ist (genauer: 0,99987). Würde es sich bei diesen Werten um die jeweilige Anzahl der Modifikationen zweier Zelllinien pro 1000 tRNAs handeln, würde dies zu einer drastischen Überschätzung der Ähnlichkeit der Zelllinien führen (obwohl z.B. bei der ersten „Modifikation“ ein Unterschied von 100% besteht und dieser Unterschied bei den realen Messergebnissen nur selten erreicht wird). Deshalb wurde jede Modifikation zunächst auf eins normiert, so dass die Summe der Quadrate der normierten Werte einer bestimmten Modifikation über alle Zelllinien eins ergibt (siehe Abschnitt 4.3.6.2). Jede Zelllinie wird auf diese Weise durch eine Zahlenreihe bzw. einen Vektor repräsentiert, dessen Komponenten durch die normierten Modifikationslevel dieser Zelllinie gebildet werden. (Die Länge dieses „Zelllinienvektors“ ist natürlich *nicht* eins.)

4.3.4.3 Durchführung der Clusteranalyse

Es wurde nun eine hierarchische Clusteranalyse mit Hilfe des von Michael Eisen entwickelten Programms *Cluster* (© Stanford University) durchgeführt.^[512] Aufgrund der gewählten Einstellungen (siehe 4.3.6.3) wurde beim Clustering zunächst dasjenige Zelllinienpaar mit dem größten Pearson-Korrelationskoeffizienten ermittelt. Aus den Vektoren, welche diese Zelllinien repräsentieren, wurde

dann ein neuer Vektor gebildet. Dieser repräsentiert dann beide Zelllinien gleichzeitig. (Man kann sich diesen Vektor als „Pseudozelllinie“ mit jeweils gemittelten Werten für die jeweiligen Modifikationen vorstellen.) Rechnerisch ergeben sich die Komponenten des Vektors aus dem Mittelwert der entsprechenden Komponenten der Ursprungsvektoren. Anschließend wird das Clustering mit diesem Vektor fortgesetzt und das nächstähnliche Vektorenpaar ermittelt. Zu beachten ist noch, dass unter den gewählten Einstellungen (siehe 4.3.6.3) bei der Berechnung eines mehrere Zelllinien repräsentierenden Vektors dessen Komponenten aus dem Mittelwert der Komponenten *der einzelnen Zelllinien* gebildet werden. Die Komponenten dieses Vektors werden also *nicht* durch Mittelung der Komponenten der beiden zuletzt zusammengefassten Vektoren gebildet (von denen einer eventuell drei Zelllinien, der andere aber nur eine Zelllinie repräsentiert).

Das Ergebnis des Clusterings ist ein hierarchisch geordnetes Dendrogramm mit den einzelnen Zelllinien als Endpunkten. Die Länge der Äste des Dendrogramms ergibt sich aus den berechneten Pearson-Korrelationskoeffizienten r durch $1-r$, d.h. je ähnlicher die Modifikationssets, umso kürzer die verbindenden Äste. Knotenpunkte (d.h. Verzweigungen) im Dendrogramm repräsentieren alle untergeordneten Zelllinien.

4.3.4.4 Ergebnisse der Clusteranalyse

In Abbildung 24 sind die Zelllinien nach Ähnlichkeit ihrer tRNA-Modifikationssets geclustert worden. Um ein fehlerhaftes Clustering zu vermeiden, wurden vier einzelne Messwerte mit zu großer Standardabweichung ($\geq 35\%$) bei diesem Clustering nicht berücksichtigt. Die Clusterergebnisse wurden mit dem ebenfalls von Michael Eisen verfassten Programm *TreeView* (© Stanford University) graphisch dargestellt.^[513]

Direkt über den Ästen des Dendrogramms ist der jeweilige Pearson-Korrelationskoeffizient r aufgeführt. Üblicherweise betrachtet man eine Korrelation nach Pearson als sehr hoch, wenn $r > 0,9$. Liegt der Korrelationskoeffizient zwischen 0,7 und 0,9, spricht man immer noch von einer hohen Korrelation. Die Korrelation ist gering, wenn r ungefähr zwischen 0,3 und 0,6 bis 0,7 liegt. Diese Festlegungen sollen auch als Grundlage für die Interpretation und Bewertung des hier durchgeführten Clusterings dienen.

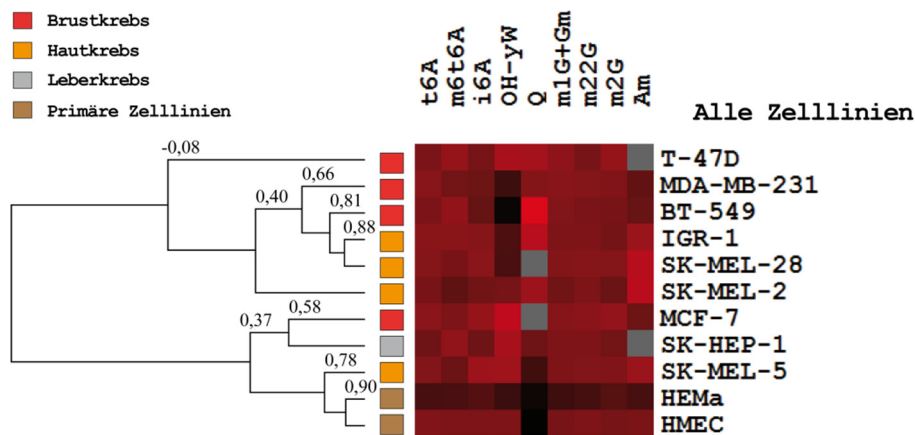


Abbildung 24. Mit Hilfe der hierarchischen Clusteranalyse auf Basis der Pearson-Korrelation generiertes Dendrogramm, dass die Ähnlichkeitsverhältnisse zwischen den tRNA-Modifikationssets der untersuchten Zelllinien veranschaulicht. In die hier dargestellte Analyse wurden alle Zelllinien und alle im Rahmen dieser Studie quantifizierten Modifikationen einbezogen. Die Länge der Äste ergibt sich durch $1-r$, wobei r der Pearson-Korrelationskoeffizient ist. Dieser ist ein Maß für die Ähnlichkeit bzw. Korrelation zweier Datenreihen ($r = 1$: Exakte Übereinstimmung; $r = 0$: Keinerlei Korrelation). Die Korrelationskoeffizienten sind im Dendrogramm aufgeführt. Die Modifikationsdichte jeder Modifikation einer jeden Zelllinie ist farbcodiert in der rechten Bildhälfte aufgeführt (schwarz: Modifikationsdichte null; hellrot: Höchste Modifikationsdichte; grau: Nichtberücksichtigte Werte).

Es fällt zunächst auf, dass sich alle Zelllinien mit Ausnahme der Brustkrebszelllinie T-47D auf zwei Cluster verteilen, welche nicht miteinander korreliert sind. Im ersten Cluster findet man zwei Brust- und drei Hautkrebszelllinien. Sehr hoch ist hier die Korrelation zwischen den Hautkrebszelllinien IGR-1 und SK-MEL-28 ($r = 0,88$). Allerdings ist schon die zu diesen beiden Proben nächstähnliche Zelllinie BT-549 keine Haut-, sondern eine Brustkrebszelllinie. Trotz des unterschiedlichen Herkunftsgewebes ist die Korrelation zwischen den drei Zelllinien mit $r = 0,88$ immer noch hoch. Die nächsthöhere Clusterstufe mit $r = 0,66$ besteht aus den drei oben genannten Zelllinien und der Brustkrebszelllinie MDA-MB-231. Schließlich findet man noch die Hautkrebszelllinie Sk-MEL-2 im ersten großen Cluster. Sie weist mit $r = 0,40$ nur noch eine geringe Korrelation zu den anderen vier Zelllinien des Clusters auf.

Der zweite große Cluster besteht auch aus fünf Zelllinien und ist noch heterogener als der erste Cluster aufgebaut. Hier finden sich die beiden primären Zelllinien, die Leberkrebs- und je eine Haut- und eine Brustkrebszelllinie. Der Cluster besteht aus zwei Unterclustern, welche nur gering miteinander korrelieren ($r = 0,37$). Der erste Untercluster besteht aus den beiden primären Zelllinien HEMa und HMEC, welche eine sehr hohe Ähnlichkeit ($r = 0,90$) aufweisen, und der Hautkrebszelllinie Sk-MEL-5. Diese korreliert immer noch stark mit den primären Zelllinien ($r = 0,78$). Der zweite

Untercluster wird von der Brustkrebszelllinie MCF-7 und der Leberkrebszelllinie SK-HEP-1 gebildet, welche relativ gering miteinander korreliert sind ($r = 0,58$).

Die Brustkrebszelllinie T-47D lässt sich keinem der beiden großen Cluster zuordnen und steht unkorreliert für sich alleine.

Während die beiden primären Zelllinien und auch zwei Hautkrebszelllinien sehr ähnliche tRNA-Modifikationssets aufweisen, hängt die gegenseitige Ähnlichkeit der Modifikationssets bei den meisten Krebszelllinien offenbar nicht sehr stark vom Herkunftsgewebe ab. Dies zeigt sich auch, wenn die Brust- und Hautkrebszelllinien jeweils für sich alleine bzw. jeweils zusammen mit den primären Zelllinien geclustert werden (Abbildung 25).

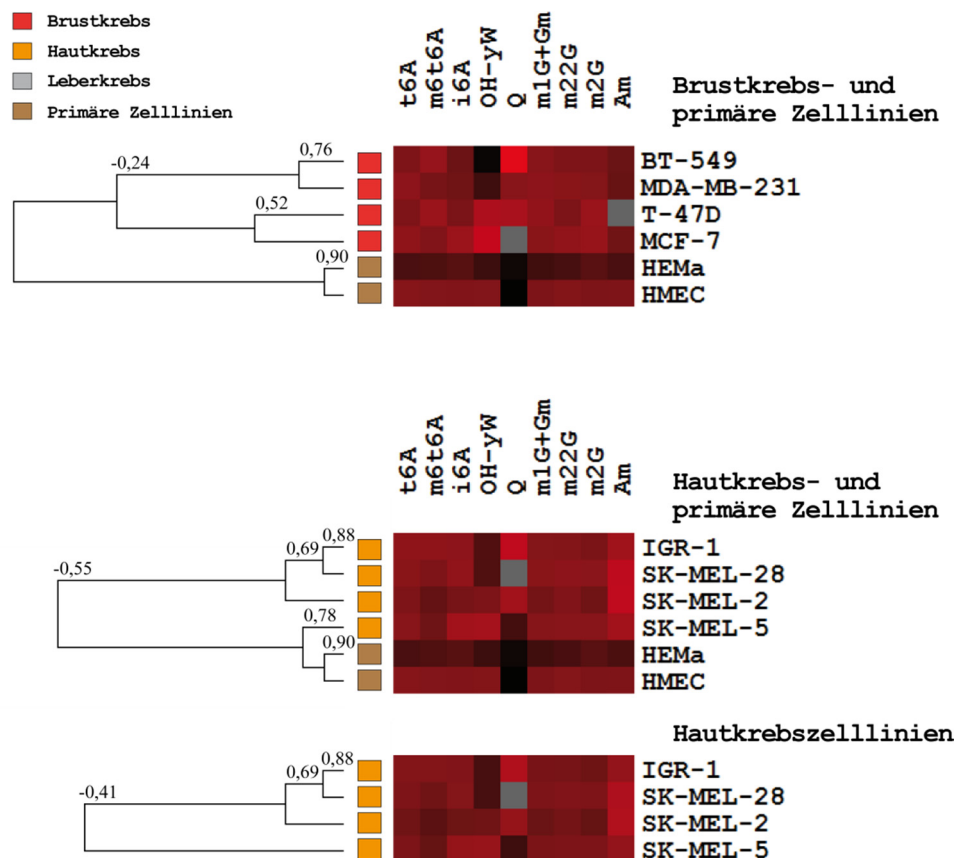


Abbildung 25. Mit Hilfe der hierarchischen Clusteranalyse berechnete Dendrogramme, welche die Ähnlichkeitsverhältnisse zwischen den tRNA-Modifikationssets der Zelllinien veranschaulichen. In den hier dargestellten Clusteranalysen wurden die Brust- (oben) bzw. Hautkrebszelllinien (Mitte) zusammen mit den primären Zelllinien und die Hautkrebszelllinien für sich alleine (unten) geclustert. Dabei wurden alle im Rahmen dieser Studie quantifizierten Modifikationen einbezogen. Die Länge der Äste ergibt sich durch $1-r$, wobei r der Pearson-Korrelationskoeffizient ist ($r = 1$: Exakte Übereinstimmung; $r = 0$: Keinerlei Korrelation). Die Korrelationskoeffizienten sind im Dendrogramm aufgeführt. Die jeweilige Modifikationsdichte ist farbcodiert in der rechten Bildhälfte angegeben (schwarz: Modifikationsdichte null; hellrot: Höchste Modifikationsdichte; grau: Nichtberücksichtigte Werte).

Eine Clusteranalyse der vier Brustkrebszelllinien zusammen mit den beiden primären Zelllinien (Abbildung 25, oben) zeigt, dass die beiden primären Zelllinien am stärksten miteinander korrelieren ($r = 0,90$). Der von diesen beiden Zelllinien gebildete Cluster steht unkorreliert neben den Brustkrebszelllinien, welche sich wiederum auf zwei nicht miteinander korrelierte Cluster aufteilen. Einer dieser Cluster wird von BT-549 und MDA-MB-231 gebildet, deren Modifikationssets eine hohe Ähnlichkeit aufweisen ($r = 0,76$). Der andere Cluster mit $r = 0,52$ zeigt, dass sich die Modifikationssets von T-47D und MCF-7 schon relativ stark unterscheiden.

Eine völlig andere Situation findet man bei der gemeinsamen Clusteranalyse von primären Zelllinien und Hautkrebszelllinien (Abbildung 25, Mitte). Hier beobachtet man zwei nicht miteinander korrelierte Cluster. Allerdings besteht der eine Cluster nicht nur aus den beiden stark korrelierten primären Zelllinien ($r = 0,90$), sondern enthält auf der nächsthöheren Clusterebene ($r = 0,78$) auch die Hautkrebszelllinie SK-MEL-5. Diesen Cluster findet man auch im Clustering aller Zelllinien als Untercluster. Die drei weiteren Hautkrebszelllinien bilden einen zweiten stark korrelierten Cluster ($r = 0,69$), welcher auch die besonders ähnlichen Zelllinien IGR-1 und SK-MEL-28 enthält ($r = 0,88$). Um die Relation zwischen SK-MEL-5 und den anderen Hautkrebszelllinien besser beurteilen zu können, wurde eine weitere Clusteranalyse dieser Zelllinien durchgeführt, diesmal ohne die primären Zelllinien (Abbildung 25, unten). Dabei zeigte sich, dass SK-MEL-5 tatsächlich nicht mit den übrigen Hautkrebszelllinien korreliert ($r = -0,41$).

Insgesamt zeigen die hier durchgeführten hierarchischen Clusteranalysen, dass sowohl innerhalb der Brust-, als auch der Hautkrebszellpopulationen beträchtliche Unterschiede bei den tRNA-Modifikationssets bestehen. Diese Unterschiede führen dazu, dass bei der gemeinsamen Clusteranalyse aller Zelllinien hauptsächlich gemischte Cluster auftreten. In diesen Clustern sind Zelllinien unterschiedlicher Herkunft z.T. stärker korreliert als Zelllinien, die aus dem gleichen Gewebetyp stammen. Die stärkste Korrelation innerhalb der zwei Krebszellpopulationen findet man bei IGR-1 und SK-MEL-28 (Haut; $r = 0,88$) bzw. bei BT-549 und MDA-MB-231 (Brust; $r = 0,76$).

Neben den Krebszelllinien bilden die beiden primären Zelllinien einen eigenen Untercluster mit einem Pearson-Korrelationskoeffizienten von $r = 0,90$. Dieser spiegelt die starke Ähnlichkeit ihrer tRNA-Modifikationssets wider und deutet darauf hin, dass sich Krebszelllinien möglicherweise anhand ihres tRNA-Modifikationssets von nicht-malignen Zelllinien unterscheiden lassen. Dagegen erscheint eine Klassifizierung von Krebszelllinien anhand ihrer tRNA-Modifikationssets nur schwer möglich.

4.3.4.5 Validierung der Clusteranalyse

In diesem Abschnitt sollen die Ergebnisse der Clusteranalyse zunächst mit literaturbekannten Clusterresultaten verglichen werden. Besonders viele Daten sind zu den sieben NCI60-Krebszelllinien verfügbar.^[475-479, 514]

Die ersten Studien, in der die NCI60-Zelllinien auf Basis ihrer Genexpressionsprofile hierarchisch geclustert worden sind, stammen von *Ross et al.*^[460] und *Scherf et al.*^[459] Von *Scherf et al.* wurden cDNA-Mikroarrays verwendet, um die Expressionsprofile von 1376 Genen parallel zu erfassen. Die Ähnlichkeit der Zelllinien wurde auch hier über die Person-Korrelation bestimmt.^[459] Es zeigte sich, dass die gefundenen Cluster oft homogen sind, d.h. hauptsächlich aus Krebszelllinien desselben Ursprungsorgans oder –gewebes bestehen. Einer dieser Cluster enthält sieben Hautkrebszelllinien (siehe Abbildung 26a, oben). Drei dieser Hautkrebszelllinien wurden auch in unserer Studie untersucht. Genau wie bei uns wurde von *Scherf et al.* eine deutlich höhere Korrelation zwischen SK-MEL-28 und SK-MEL-2 gefunden, als sie zwischen jeder dieser beiden Zelllinien und SK-MEL-5 besteht (siehe Abbildung 26a). Die Brustkrebszelllinien erweisen sich bei *Scherf et al.* als besonders heterogen und sind als einziger Krebszelllinientyp in allen großen Clustern zu finden. Bemerkenswerterweise wird genau wie bei uns eine Korrelation zwischen MCF-7 / T-47D einerseits und MDA-MB-231 / BT-549 andererseits beobachtet, während diese beiden Zelllinienpaare ebenso wie bei uns (siehe Abbildung 26b, unten) keinerlei Korrelation untereinander aufweisen. Während MCF-7 und T-47D mit den Darmkrebszelllinien clustern, findet man BT-549 im Cluster der ZNS-Krebszelllinien. Dieser ZNS-Cluster ist Teil eines größeren Clusters, welcher auch die Zelllinie MDA-MB-231 enthält.

In einer Studie von *Perou et al.* wurden 19 Brustkrebszelllinien und diverse Biopate durch hierarchisches Clustering der Expressionsprofile von 1753 Genen genauer klassifiziert.^[461] Wieder wurde der Pearson-Korrelationskoeffizient als Maß für die Ähnlichkeit zweier Genexpressionsprofile bzw. Zelllinien gewählt. Genau wie von uns und von *Scherf et al.* beobachtet,^[459] wurde auch in dieser Studie eine hohe Korrelation zwischen MCF-7 / T-47D einerseits und MDA-MB-231 / BT-549 andererseits gefunden (siehe Abbildung 26b).

Von Cravatt und Mitarbeitern wurde eine Klassifizierung von humanen Brust- und Hautkrebszelllinien auf Basis ihrer Serin-Hydrolase-Aktivitätsprofile durchgeführt.^[471] Dieses wurden zunächst über ABPP-Methode (*Activity-Based Protein Profiling*) ermittelt.^[470] Über den auf der Person-Korrelation basierenden hierarchischen Clusteralgorithmus, welcher auch von uns verwendet wurde, wurde dann die Korrelation zwischen den Zelllinien berechnet. Es zeigte sich, dass die untersuchten Zelllinien drei

Cluster bilden: Einen Hautkrebscluster, einen Brustkrebscluster und einen Cluster mit besonders invasiven Zelllinien (siehe **Abbildung 26c**).^[471] Vier der elf von Cravatt und Mitarbeitern untersuchten Zelllinien sind auch von uns untersucht worden. Dabei handelt es sich um drei Brustkrebszelllinien (T-47D, MCF-7, MDA-MB-231) und eine Hautkrebszelllinie (SK-MEL-2.). T-47D und MCF-7 bilden bei Cravatt und Mitarbeitern den Brustkrebscluster, während SK-MEL-2 mit den anderen Hautkrebszelllinien clustert. Die Zelllinie MDA-MB-231 findet man dagegen im Cluster der besonders invasiven Zelllinien. Da das uns ermittelte Verhältnis dieser Zelllinien zueinander nicht direkt aus Abbildung 24 zu entnehmen ist, wurden die Zelllinien nochmals separat geclustert (siehe Abbildung 26c, unten). Das Ergebnis dieses Clusterings stimmt qualitativ sehr gut mit den Beobachtungen von Cravatt und Mitarbeitern überein: Während die Zelllinien T-47D und MCF-7 korreliert sind (allerdings nur schwach; $r = 0,52$), stehen die Zelllinien MDA-MB-231 und SK-MEL-2 unkorreliert neben diesem Cluster ($r = -0,11$).

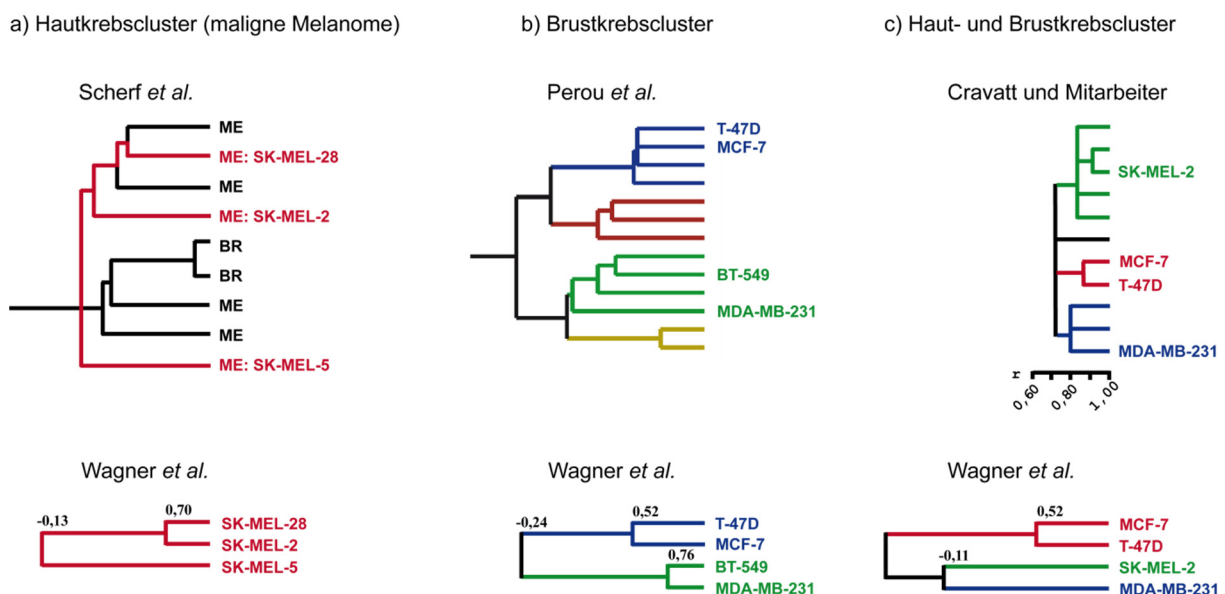


Abbildung 26. Durch hierarchische Clusteranalyse generierte Dendrogramme, welche die Ähnlichkeitsverhältnisse zwischen Krebszelllinien veranschaulichen. Die Dendrogramme in der oberen Bildhälfte stammen von anderen Arbeitsgruppen und sind literaturbekannt, darunter sind die von mir durch Clusteranalyse der entsprechenden tRNA-Modifikationssets erhaltenen Dendrogramme zu sehen. Es liegt jeweils ein unterschiedlicher Maßstab für die Länge der Äste vor. a) Oben: Darstellung des von Scherf *et al.* durch Analyse von Genexpressionsprofilen gefundenen Hautkrebsclusters.^[459] ME: (Melanom-)Hautkrebszelllinie; BR: Brustkrebszelllinie. b) Oben: Ausschnitt aus dem von Perou *et al.* durch Analyse von Genexpressionsprofilen gefundenen Brustkrebscluster.^[461] Blau: Krebszelllinien aus dem luminalen Epithel der Brust; grün: mesenchymartige Brustkrebszelllinien; rot: Krebszelllinien aus den Lymphozyten der Brust; gelb: Krebszelllinien aus dem Endothel der Brust. c) Oben: Von Cravatt und Mitarbeitern durch Clusteranalyse von Proteinaktivitätsprofilen gefundenes Dendrogramm.^[471] Grün: Melanomcluster; rot: Brustkrebscluster; blau: Cluster mit besonders invasiven Krebszelllinien.

Um die Clusterresultate dieser Studie weiter zu validieren, wurde der Einfluss einzelner tRNA-Modifikationen auf das Resultat des Clusterings systematisch untersucht. Dazu wurden bestimmte Modifikationen nacheinander beim Clustering nicht berücksichtigt. Durch den Vergleich der so erhaltenen alternativen Dendrogramme mit dem ursprünglichen Dendrogramm aus Abbildung 24 kann dann abgeleitet werden, wie stark der Einfluss einzelner Modifikationen auf das Gesamtergebnis ist. Die oben diskutierten Korrelationen zwischen den Zelllinien sind als signifikant zu betrachten, wenn sich das ursprüngliche Clustering auch bei der Nichtberücksichtigung einzelner Modifikationen kaum ändert.

Wie in Abbildung 23 zu sehen ist, findet man die größten Schwankungen der zelllinienspezifischen Modifikationsdichte bei den Modifikationen Am, HO-yW und Q. Hier betragen die prozentualen Unterschiede zwischen kleinster und größter Modifikationsdichte bis zu 220% (Q). Es war daher zu erwarten, dass sich eine Nichtberücksichtigung dieser Modifikationen am stärksten auf das Gesamtergebnis der statischen Analyse auswirken sollte. Deshalb wurden zunächst auf der Pearson-Korrelation basierende hierarchische Clusteranalysen der tRNA-Modifikationssets durchgeführt, bei denen jeweils eine dieser drei Modifikationen nicht miteinbezogen wurde. Zur Beurteilung der alternativen Ergebnisse wurden drei Pearson-Korrelationskoeffizienten als Vergleichswerte ausgewählt. Diese Werte waren beim Clustering unter Einbeziehung aller Modifikationen ermittelt worden. Es handelt sich hierbei um die Korrelationskoeffizienten der beiden primären Zelllinien (HEMA und HMEC; $r = 0,90$; Abbildung 25, oben und Mitte) und der jeweils ähnlichsten Brust- (BT-549 und MDA-MB-231; $r = 0,76$; Abbildung 25, oben) und Hautkrebszelllinien (IGR-1 und SK-MEL-28; $r = 0,88$; Abbildung 25, Mitte und unten). Die unter Nichtberücksichtigung von Am, HO-yW bzw. Q durch das hierarchische Clustering generierten alternativen Dendrogramme finden sich in den Abbildungen 9-12.

Das Clustering ohne Am führt auf den ersten Blick zu einer größeren Veränderung des Dendrogramms (Abbildung 27). Man beobachtet nun drei ungleiche Paare aus einer Brust- und einer Hautkrebszelllinie, deren Korrelation jeweils hoch bzw. sehr hoch ist (von $r = 0,83$ bis $r = 0,96$ bei BT-549 und IGR-1). Allerdings ist das alternative Clustering dem ursprünglichen Clustering bei näherer Betrachtung immer noch sehr ähnlich. Die beiden ursprünglich identifizierten, heterogenen Cluster finden sich in nur leicht veränderter Form auch in diesem Clustering wieder. Der größte Unterschied ist das Auftreten eines neuen Clusters mit geringer Korrelation ($r = 0,52$), der aus den Zelllinien SK-HEP-1 und T-47D gebildet wird. T-47D stand im ursprünglichen Clustering unkorreliert neben den anderen Zelllinien, während SK-HEP-1 ein gering korrelierter Teil des die primären Zelllinien enthaltenden Clusters war. Auch ohne Berücksichtigung von Am ändert sich die Korrelation zwischen den primären Zelllinien nicht und ist immer noch sehr hoch ($r = 0,90$). Auch die Korrelationen zwischen

den als weitere Referenzpaare gewählten Zelllinien ändern sich praktisch nicht: Bei den Brustkrebszelllinien BT-549 und MDA-MB-231 beträgt die Korrelation immer noch 0,76; bei den Hautkrebszelllinien IGR-1 und SK-MEL-28 0,87 statt 0,88. Offenbar wird das Gesamtergebnis des Clusterings also kaum von Am beeinflusst.

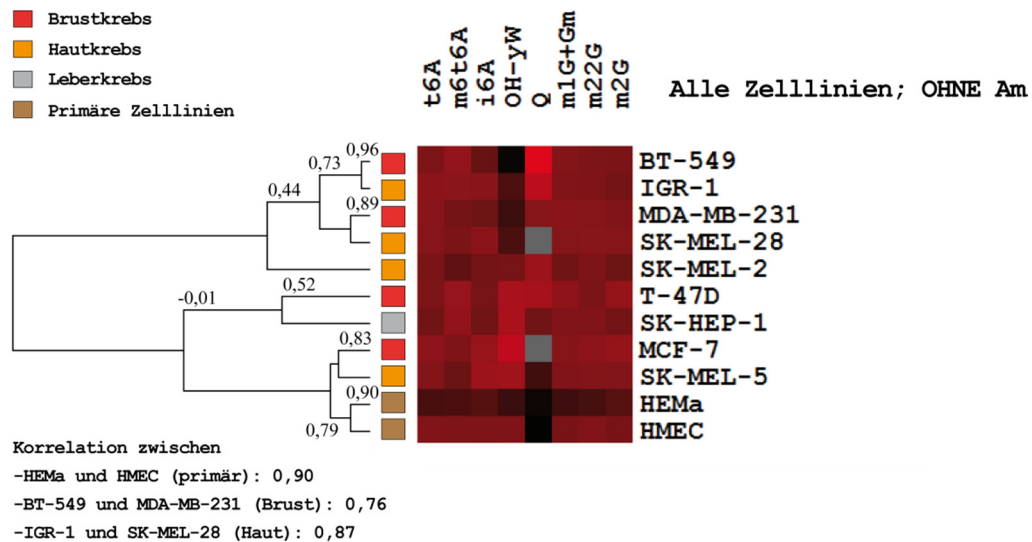


Abbildung 27. Mit Hilfe der hierarchischen Clusteranalyse berechnetes Dendrogramm, welches die Ähnlichkeitsverhältnisse zwischen den tRNA-Modifikationssets der in dieser Studie untersuchten Zelllinien veranschaulicht. Die Modifikation Am wurde bei dieser Clusteranalyse nicht berücksichtigt. Die Länge der Äste ergibt sich durch $1-r$, wobei r der Pearson-Korrelationskoeffizient ist ($r = 1$: Exakte Übereinstimmung; $r = 0$: Keinerlei Korrelation). Die Korrelationskoeffizienten sind im Dendrogramm aufgeführt. Die jeweilige Modifikationsdichte ist farbcodiert in der rechten Bildhälfte angegeben (schwarz: Modifikationsdichte null; hellrot: Höchste Modifikationsdichte; grau: Nichtberücksichtigte Werte).

Bei der nächsten Clusteranalyse wurde die Modifikation HO-yW nicht berücksichtigt (**Abbildung 28**). Das so erhaltene Dendrogramm entspricht sogar noch mehr unseren ursprünglichen Erwartungen. Es gibt nun drei große Cluster, von denen zwei homogen aufgebaut sind und nur Zelllinien des gleichen Herkunftsgewebes enthalten. Der erste Cluster enthält alle Brustkrebszelllinien, während der zweite Cluster aus drei der vier Hautkrebszelllinien besteht. Der dritte Cluster enthält neben den sehr stark korrelierten primären Zelllinien ($r = 0,94$) wieder die Hautkrebszelllinie SK-MEL-5. Diese Zelllinie ist schon zuvor durch eine starke Korrelation mit den primären Zelllinien aufgefallen und hat in keinem Clustering eine Ähnlichkeit zu den anderen Hautkrebszelllinien aufgewiesen. Die Leberkrebszelllinie SK-HEP-1 findet man bei diesem Clustering unkorreliert neben den drei großen Clustern stehend. Insgesamt zeigt die Clusteranalyse ohne HO-yW, dass die tRNA-Sets von Krebszelllinien aus dem gleichen Herkunftsgewebe möglicherweise eine größere Ähnlichkeit aufweisen, als zunächst erwartet.

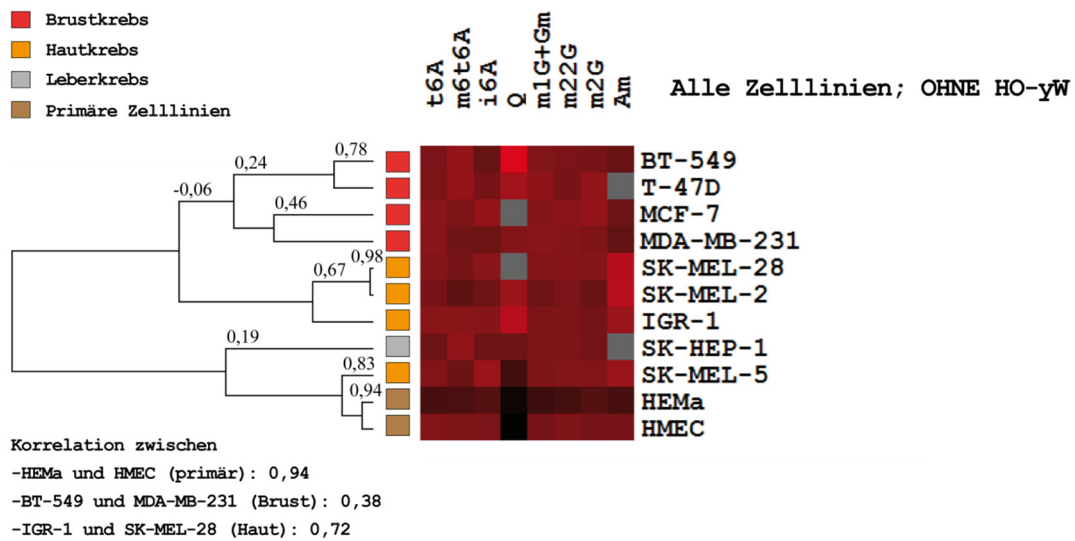


Abbildung 28. Mit Hilfe der hierarchischen Clusteranalyse berechnetes Dendrogramm, welches die Ähnlichkeitsverhältnisse zwischen den tRNA-Modifikationssets der in dieser Studie untersuchten Zelllinien veranschaulicht. Die Modifikation HO-yW wurde bei dieser Clusteranalyse nicht berücksichtigt. Die Länge der Äste ergibt sich durch $1-r$, wobei r der Pearson-Korrelationskoeffizient ist ($r = 1$: Exakte Übereinstimmung; $r = 0$: Keinerlei Korrelation). Die Korrelationskoeffizienten sind im Dendrogramm aufgeführt. Die jeweilige Modifikationsdichte ist farbcodiert in der rechten Bildhälfte angegeben (schwarz: Modifikationsdichte null; hellrot: Höchste Modifikationsdichte; grau: Nichtberücksichtigte Werte).

Eine Clusteranalyse ohne die Modifikation Q führt dagegen zu einigen wesentlichen Veränderungen (**Abbildung 29**). Man beobachtet nun zwei Cluster auf höchster Ebene, neben denen die Hautkrebszelllinie SK-MEL-5, aber auch die primäre Zelllinie HMEC unkorreliert stehen. Der kleinere Cluster besteht nur aus den Brustkrebszelllinien T47-D und MCF-7 und der Leberkrebszelllinie SK-HEP-1. Im zweiten Cluster sind die restlichen Zelllinien, einschließlich der zweiten primären Zelllinie HEMa, vereinigt. Obwohl sich diese Verteilung stark vom ursprünglichen Clustering unterscheidet, weisen die als Referenzpaare gewählten Krebszelllinien (BT-549 und MDA-MB-231; Brust; $r = 0,76$; IGR-1 und SK-MEL-28; Haut; $r = 0,88$) mit $r = 0,88$ bei BT-549 / MDA-MB-231 bzw. $r = 0,88$ bei IGR-1 / SK-MEL-28 wie zuvor eine sehr hohe Korrelation auf. Allerdings führt die Nichtberücksichtigung von Q bei der Clusteranalyse dazu, dass die beiden primären Zelllinien überraschenderweise keinerlei Korrelation mehr aufweisen ($r = -0,22$).

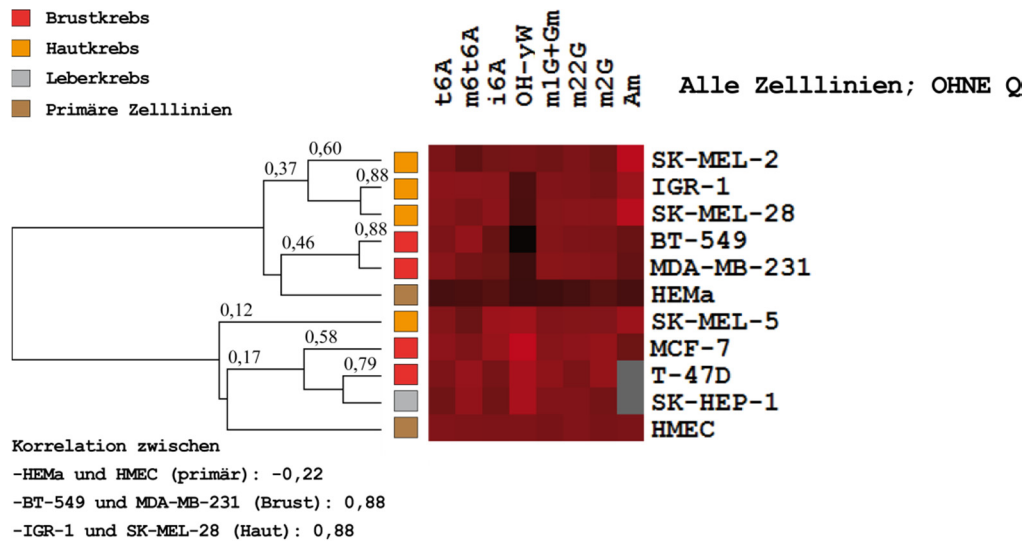


Abbildung 29. Mit Hilfe der hierarchischen Clusteranalyse berechnetes Dendrogramm, welches die Ähnlichkeitsverhältnisse zwischen den tRNA-Modifikationssets der in dieser Studie untersuchten Zelllinien veranschaulicht. Die Modifikation Q wurde bei dieser Clusteranalyse nicht berücksichtigt. Die Länge der Äste ergibt sich durch $1-r$, wobei r der Pearson-Korrelationskoeffizient ist ($r = 1$: Exakte Übereinstimmung; $r = 0$: Keinerlei Korrelation). Die Korrelationskoeffizienten sind im Dendrogramm aufgeführt. Die jeweilige Modifikationsdichte ist farbcodiert in der rechten Bildhälfte angegeben (schwarz: Modifikationsdichte null; hellrot: Höchste Modifikationsdichte; grau: Nichtberücksichtigte Werte).

Abschließend wurden drei weitere Analysen zur Verifikation der ursprünglichen Clusterergebnisse durchgeführt, bei denen jeweils eine bestimmte Modifikation nicht berücksichtigt wurde. Nun allerdings wurden nacheinander m^6t^6A , t^6A und m^2G von der Analyse ausgenommen, welche alle nur geringe Schwankungen der Modifikationsdichte zwischen den Zelllinien aufweisen.

Die Ergebnisse dieser Clusteranalysen sind in Abbildung 30 dargestellt. Wie erwartet, ist der Einfluss jeder dieser Modifikationen für sich auf das Gesamtergebnis des Clusterings gering. Dies zeigt sich besonders an den hier erhaltenen Dendrogrammen, welche dem ursprünglichen Dendrogramm außerordentlich ähnlich sind. Außerdem weisen die Referenzzelllinienpaare in allen drei Analysen eine hohe Korrelation auf, wie sie zuvor auch beim Clustering unter Einbeziehung aller quantifizierten Modifikationen beobachtet worden ist.

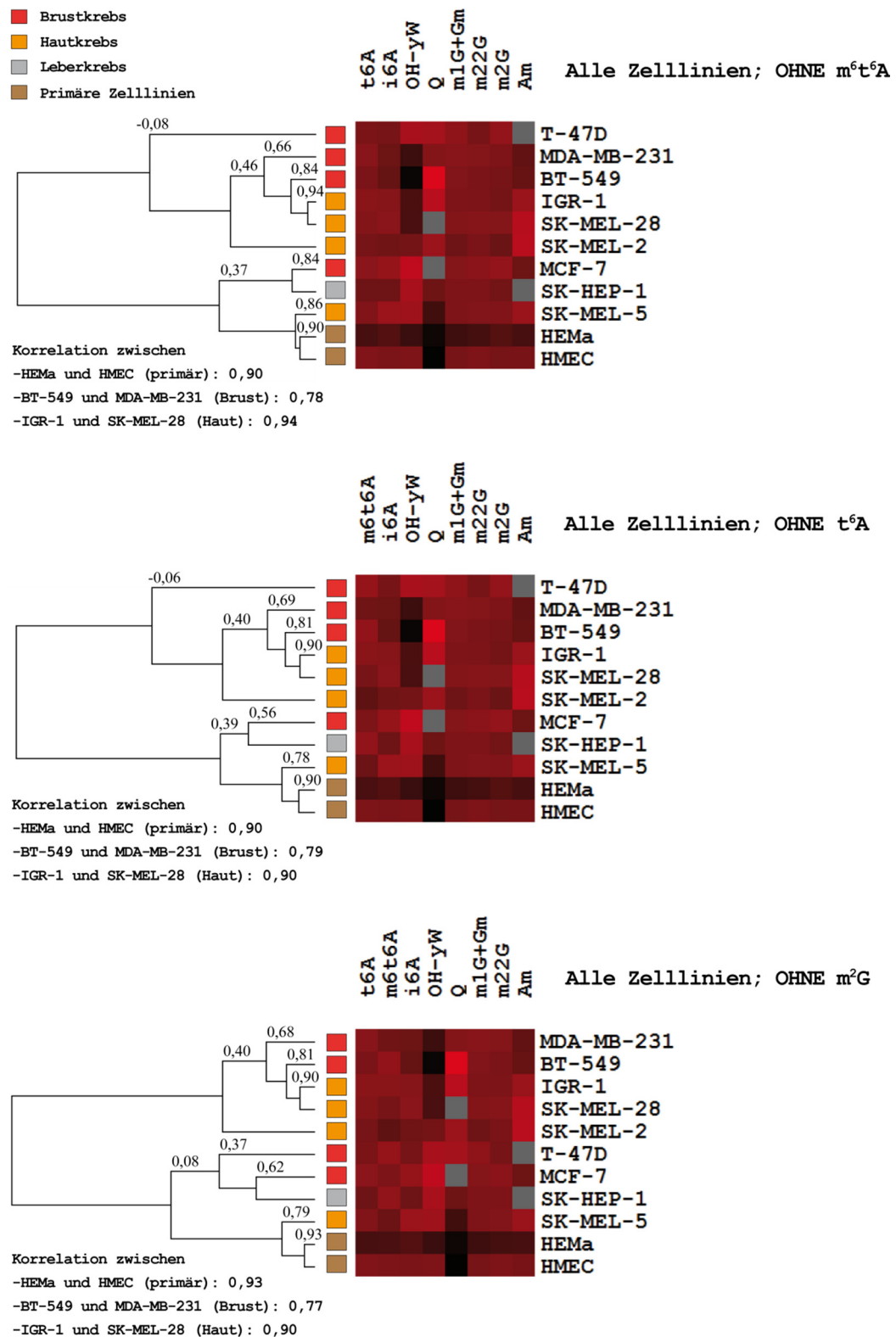


Abbildung 30. Mit Hilfe der hierarchischen Clusteranalyse berechnete Dendrogramme, welche die Ähnlichkeitsverhältnisse zwischen den tRNA-Modifikationssets der Zelllinien veranschaulichen. Die Modifikation m⁶t⁶A wurde für die Berechnung des oberen Dendrogramms nicht berücksichtigt, während t⁶A bzw. m²G für die Berechnung des mittleren bzw. unteren Dendrogramms nicht berücksichtigt wurden.

4.3.5 Zusammenfassung und Ausblick

Wie in dieser Studie gezeigt werden konnte, weisen alle untersuchten Krebszelllinien ein individuelles Set an tRNA-Modifikationen auf, das sich von den Modifikationssets anderer Zelllinien z.T. deutlich unterscheidet. Daher lassen sich bekannte Krebszelllinien anhand ihrer individuellen tRNA-Modifikationssets ähnlich wie mit einem Fingerabdruck zu identifizieren. Auch die hier untersuchten primären Zelllinien besitzen jeweils individuelle Modifikationssets.

Die auf Basis der Pearson-Korrelation^[506] durchgeführten hierarchischen Clusteranalysen^[510, 515] haben gezeigt, dass sich die Modifikationssets der beiden primären Zelllinien sehr ähnlich sind. Sie lassen sich daher über ihre tRNA-Modifikationssets von den Krebszelllinien unterscheiden. Interessanterweise scheint der wesentliche Unterschied zwischen primären und Krebszelllinien darin zu bestehen, dass primäre Zelllinien geringere Schwankungen in der relativen Dichte der einzelnen Modifikationen aufweisen. Dagegen findet man den hier untersuchten Krebszelllinien jeweils eine oder mehrere tRNA-Modifikationen, die überdurchschnittlich hoher bzw. geringer Menge in dieser Zelllinie auftreten.

Weiterhin haben wir auf Basis der von uns quantifizierten Modifikationen beobachtet, dass sich die tRNA-Modifikationssets einer Brustkrebs- und einer Hautkrebszelllinie in einigen Fällen ähnlicher sind, als z.B. die Modifikationssets zweier Brustkrebszelllinien. In guter Übereinstimmung dazu haben Genexpressionsanalysen und darauf basierende hierarchische Clusterings gezeigt, dass Brustkrebszelllinien im Vergleich zu anderen Krebsarten eine außergewöhnliche Variationsbreite aufweisen.^[459-460] Diese Variationsbreite führt dazu, dass die Brustkrebszelllinien nicht untereinander, sondern mit verschiedenen anderen Krebszelllinien clustern.

Es wurde außerdem von uns gezeigt, dass die Art und Anzahl der in dieser Studie quantifizierten modifizierten tRNA-Nukleoside ausreicht, um Ähnlichkeiten zwischen bestimmten tRNA-Modifikationssets und damit zwischen den entsprechenden Zelllinien verlässlich zu identifizieren. Die Auswahl an Modifikationen ist so robust, dass Clusteranalysen unter Nichtberücksichtigung einzelner Modifikationen immer noch ähnliche Clusterergebnisse liefern. Außerdem stimmen die von uns ermittelten Korrelationen zwischen den Zelllinien qualitativ sehr gut mit literaturbekannten Korrelationen überein, welche anhand von Genexpressionsanalysen^[459, 461] oder Enzymaktivitätsprofilen^[471] bestimmt wurden.

Für die Zukunft ist es wünschenswert, die Studie auf weitere primäre Zelllinien auszudehnen und dabei weitere tRNA-Modifikationen zu berücksichtigen. Dann wird sich der Unterschied zwischen primären und Krebszelllinien mit noch größerer statistischer Signifikanz als bisher zeigen lassen.

Das Clustering ohne Berücksichtigung von HO-yW lässt es sogar möglich erscheinen, dass bei einer statistischen Analyse auf Basis einer noch größeren Anzahl an tRNA-Modifikationen Krebszelllinien ihrem Ursprungsgewebe (d.h. Brust, Haut etc.) zugeordnet werden könnten.

4.3.6 Anhang; experimenteller Teil

4.3.6.1 Zellkultur, Extraktion der tRNA, enzymatischer Verdau und LC-MS-Messungen

Die Kultivierung der Krebszelllinien, die Extraktion und Aufreinigung der tRNA-Sets aus diesen Zelllinien, ihr enzymatischer Verdau zu den Nukleosiden und die massenspektrometrischen Quantifizierungen sind bereits an anderer Stelle publiziert und dort genau beschrieben worden (siehe dazu auch Kapitel 4.1 und 4.2.).^[97, 100]

4.3.6.2 Normierung der Nukleosidlevel für die Clusteranalysen

Die Grundlage der gesamten hier durchgeführten statistischen Clusteranalysen sind die LC-MS-Quantifizierungsdaten der tRNA-Modifikationen. Die im Arbeitskreis entwickelte und bereits publizierte Isotopenverdünnungsmethode erlaubte mir eine exakte absolute Bestimmung der in den Proben vorhandenen Stoffmengen. Da die durchschnittliche molare Masse eines tRNA-Moleküls und die Gesamtmasse der analysierten tRNA bekannt sind, lassen sich die ermittelten Stoffmengen der Nukleoside leicht in *Modifikationen pro 1000 tRNAs*, d.h. in Promille umrechnen. In Tabelle S1 sind diese Promillewerte für alle Zelllinien und Modifikationen aufgeführt.

Tabelle S1. In dieser Studie untersuchte tRNA-Modifikationen der Krebszelllinien, angegeben in *Modifikationen pro 1000 tRNAs*.

%o tRNA	t6A	m6t6A	i6A	OH-yW	Q	m1G+Gm	m22G	m2G	Am
MDA-MB-231	272	18	37	10	62	898	579	992	14
MCF-7	278	19	50	28	65	879	608	1150	15
T-47D	247	23	40	25	78	941	526	1154	26
BT-549	247	23	35	3	105	862	538	938	15
SK-MEL-2	242	15	39	18	74	755	540	848	26
SK-MEL-5	260	17	52	24	33	844	568	1003	21
SK-MEL-28	264	19	46	12	51	870	586	1026	25
IGR-1	276	21	45	12	88	839	549	909	21
SK-HEP-1	228	22	38	25	54	843	556	901	26
HMEC	257	19	41	19	4	794	554	929	17
HEMa	153	12	29	9	12	450	326	689	11

Um mit den bei meinen Messungen erhaltenen Daten eine Clusteranalyse durchführen zu können, war zunächst eine Normierung der für die verschiedenen Modifikationen erhaltenen Messwerte notwendig (siehe dazu auch Abschnitt 4.3.4.2). Deshalb wurde jede Modifikation zunächst auf eins normiert, so dass die Summe der Quadrate der normierten Werte einer bestimmten Modifikation über alle Zelllinien eins ergibt. Jede Zelllinie wird auf diese Weise durch eine Zahlenreihe bzw. einen Vektor repräsentiert, dessen Komponenten durch die normierten Modifikationslevel dieser Zelllinie gebildet werden. (Die Länge dieses „Zelllinienvektors“ ist natürlich *nicht* eins.) Die normierten Werte für jede Modifikation einer jeden Zelllinie sind in Tabelle S2 aufgeführt. Dieser Datensatz kann in das Statistikprogramm Cluster eingelesen und dann ausgewertet werden. Einzelheiten dazu finden sich im *Cluster and TreeView Manual* von Michael Eisen.^[516]

Tabelle S2. Für die statistische Clusteranalyse auf 1 normierte Werte der tRNA-Modifikationen. Die Summe der Quadrate der hier aufgeführten Werte einer bestimmten Modifikation über alle Zelllinien, d.h. in einer Spalte, ergibt eins.

tRNA (normiert)	t6A	m6t6A	i6A	OH-yW	Q	m1G+Gm	m22G	m2G	Am
MDA-MB-231	0,328	0,280	0,267	0,159	0,293	0,328	0,321	0,309	0,207
MCF-7	0,335	0,303	0,364	0,466	0,307	0,321	0,337	0,359	0,225
T-47D	0,298	0,362	0,288	0,407	0,366	0,344	0,292	0,360	0,379
BT-549	0,298	0,354	0,253	0,049	0,494	0,315	0,298	0,293	0,216
SK-MEL-2	0,292	0,238	0,279	0,288	0,350	0,276	0,300	0,265	0,379
SK-MEL-5	0,314	0,263	0,379	0,386	0,156	0,308	0,315	0,313	0,316
SK-MEL-28	0,319	0,293	0,335	0,193	0,239	0,318	0,325	0,320	0,374
IGR-1	0,333	0,332	0,325	0,196	0,413	0,306	0,304	0,284	0,315
SK-HEP-1	0,275	0,350	0,274	0,407	0,252	0,308	0,308	0,281	0,384
HMEC	0,311	0,303	0,300	0,305	0,019	0,290	0,307	0,290	0,248
HEMa	0,185	0,193	0,212	0,154	0,054	0,164	0,181	0,215	0,157

4.3.6.3 Für die Clusteranalyse der Modifikationssets der Krebszelllinien in *Cluster* gewählte Programmeinstellungen

Eine hierarchische Clusteranalyse mit diesen normierten Werten wurde von mir mit Hilfe des von Michael Eisen entwickelten Programms *Cluster* (© Stanford University) durchgeführt.^[517] Es wurden dabei folgende Einstellungen gewählt:

Filter Data-Funktion: War bei dieser Analyse nicht notwendig.

Adjust Data-Funktion: War bei dieser Analyse auch nicht notwendig.

Das Programm bietet dann die Möglichkeit, aus den folgenden Clusteralgorithmen eine Auswahl zu treffen: *Hierarchical Clustering*; *K Means Clustering*; *Self Organizing Maps*; *PCA*. Hier ist das *Hierarchical Clustering* die beste Wahl. Unter diesem Menüpunkt habe ich dann die folgenden Einstellungen vorgenommen:

1)

Genes:

Cluster – YES; Calculate Weights - NO

Arrays:

Cluster – NO; Calculate Weights - NO

Genes /Similarity Metric: Correlation (uncentered)

2)

Average Linkage Clustering anwählen.

Das Programm berechnet dann die Ähnlichkeiten zwischen den eingegebenen Einzeldatensätzen bzw. Vektoren (hier: den tRNA-Modifikationssets der Krebszelllinien). Das Ergebnis des Clusterings wird als Datenfile ausgegeben, welches mit dem Programm *TreeView* geöffnet werden kann. *TreeView* ist das von Eisen *et al.* entwickelte Programm zur grafischen Darstellung des bei der Clusteranalyse gefundenen Dendrogramms, in dem die Einzeldatensätze nach ihrer Ähnlichkeit angeordnet sind (siehe oben).

4.4 “Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain”

M. Münzel, D. Globisch, T. Brückl, **M. Wagner**, V. Welzmler, S. Michalak, M. Müller, M. Biel, T. Carell, *Angew. Chem., Int. Ed. Engl.* **2010**, 49, 5375-5377.

2009 war TET1 als zuvor unbekanntes DNA-modifizierendes Enzym identifiziert worden, welches durch Oxidation von mC in Säugetieren die Base 5-Hydroxymethylcytosin bilden kann. In den ersten beiden Publikationen wurde gezeigt, dass die höchsten hmC-Level im Gehirn bzw. in Neuronen zu finden sind. Allerdings waren die dort publizierten hmC-Level aufgrund der verwendeten, Dünnschichtchromatographie-basierten Quantifizierungsmethode noch sehr ungenau. Ziel dieses Projektes war es daher, die von uns für die Quantifizierung von RNA-Nukleosiden entwickelte LC-MS-basierte Isotopenverdünnungsmethode für die exakte und parallele Quantifizierung von DNA-Modifikationen zu adaptieren. Für eine erste Demonstration der Leistungsfähigkeit der adaptierten Quantifizierungsmethode sollte wegen der zuvor entdeckten hohen hmC-Level in Neuronen eine Analyse der verschiedenen Hirnareale der Maus durchgeführt werden.

In dieser Publikation wurde das adaptierte Quantifizierungsverfahren und die Synthese der für dieses Verfahren benötigten schweratommarkierten DNA-Nukleoside erstmals vorgestellt. Mit dieser Quantifizierungsmethode konnten wir die exakten hmC- und mC-Level in verschiedenen Hirngeweben bestimmen. Es zeigte sich, dass das hmC-Level zwischen den untersuchten Hirngeweben stark variiert, während in allen Geweben das gleiche mC-Level detektiert wurde. Weiterhin wurde von uns im Hippocampus einen Anstieg des hmC-Levels in Abhängigkeit vom Lebensalter beobachtet. Bei der Untersuchung einer neuronalen und einer glialen Krebszelllinie konnte kein hmC nachgewiesen werden.

Für dieses Projekt habe ich die neuronale murine Krebszelllinie *Neuro-2a* und die gliale humane Krebszelllinie *U-87 MG* in Kultur genommen, ihre DNA isoliert und für die Messungen zur Verfügung gestellt. Außerdem habe ich auch zusammen mit Veronika Reiter, geb. Welzmler, aus allen Gewebeproben die für die Messungen benötigte DNA isoliert.

Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain**

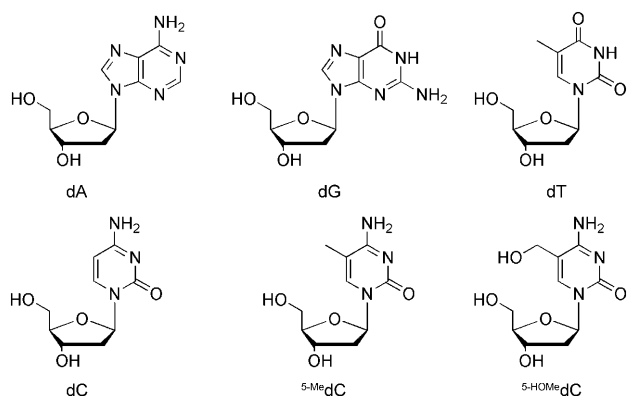
Martin Münzel, Daniel Globisch, Tobias Brückl, Mirko Wagner, Veronika Welzmler, Stylianos Michalakos, Markus Müller, Martin Biel, and Thomas Carell*

The genetic code is established by the sequence of the four canonical DNA nucleosides dA, dC, dG, and dT.^[1] Of these four bases, only the dC base is chemically modified inside cells of higher organisms to control transcriptional activity.^[2] Special methyltransferases replace the H atom at position 5 by a methyl group to form methylcytosine (⁵-Me dC).^[3] Methylation occurs only in CpG sequences and is mostly responsible for the silencing of genes.^[4] In two recent publications 5-hydroxymethylcytosine (⁵-HOMe dC) was established as a new post-replicatively formed DNA nucleoside (Scheme 1). Kriaucionis and Heintz detected ⁵-HOMe dC in cerebellar purkinje neurons.^[5] Tahiliani et al. reported the presence of traces of ⁵-HOMe dC (ca. 0.032 % of all nucleosides)

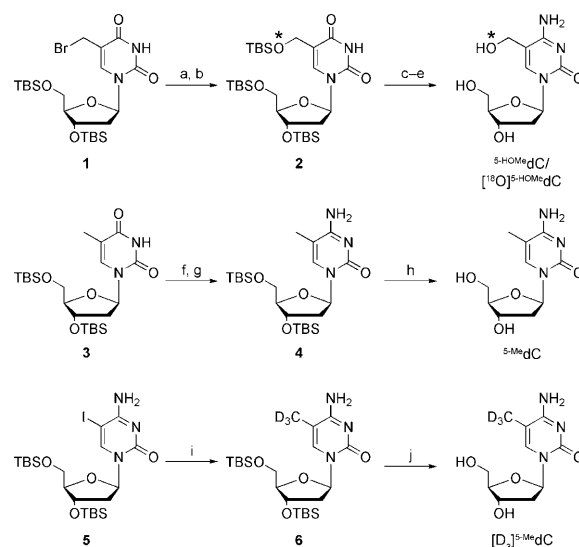
in mouse embryonic stem cells and observed the new base in CpG sequences in human embryonic kidney (HEK) cells when they overexpressed the hydroxylating 2-oxoglutarate and Fe^{II}-dependent enzyme TET in these cells.^[6] It was additionally shown that these TET enzymes are able to oxidize the 5-methyl group of ⁵-Me dC to give the hydroxymethyl group in vitro. The new base was detected using thin-layer chromatography after radioactive labeling of the nucleotides. The function of ⁵-HOMe dC is currently unclear, but it is speculated that it may establish another level of transcriptional control or that it is an intermediate of a putative oxidative demethylation mechanism.^[7]

We developed a quantitative LC-MS method to investigate the distribution of ⁵-HOMe dC in mammal brains and to determine the relative quantities of ⁵-HOMe dC and ⁵-Me dC. To this end, we synthesized both nucleosides in natural and isotope-labeled forms^[8] (Scheme 2) and quantified their amounts in different mouse brain tissues.

⁵-HOMe dC and ⁵-Me dC were labeled as ¹⁸O and CD₃ derivatives, respectively. The synthesis of ⁵-HOMe dC started



Scheme 1. Structure of the four canonical nucleosides and of the post-replicatively formed bases ⁵-Me dC and ⁵-HOMe dC.



Scheme 2. Syntheses of the isotope-labeled nucleosides ⁵-HOMe dC, [¹⁸O]⁵-HOMe dC, ⁵-Me dC, and [D₃]⁵-Me dC. a) H₂O, DIPEA, 72 %, b) TBSCl, imid., 46 %, c) NaH, TPSCl, d) NH₃/MeOH, 76 % over two steps, e) 3 HF·NET₃, 52 %, f) NaH, TPSCl, g) NH₃/MeOH, 81 % over two steps, h) HF-pyr, 88 %, i) CD₃MgI, CuCl, [Pd(PPh₃)₄], 90 % inseparable mixture of **6** and TBSdC, j) HF-pyr, 84 % (based on pure **6**). [¹⁸O]⁵-HOMe dC was synthesized by the same route as ⁵-HOMe dC. The asterisk (*) indicates the ¹⁸O label. DIPEA = N,N-diisopropylethylamine, imid. = imidazole, pyr = pyridine, TBS = *tert*-butyldimethylsilyl, TPS = 2,4,6-triisopropylbenzenesulfonyl.

[*] Dipl.-Chem. M. Münzel,^[+] Dipl.-Chem. D. Globisch,^[+] Dipl.-Chem. T. Brückl, Dipl.-Chem. M. Wagner, Dipl.-Chem. V. Welzmler, Dr. M. Müller, Prof. Dr. T. Carell Center for Integrated Protein Science (CiPS^M) at the Department of Chemistry, Ludwig-Maximilians-University Butenandtstrasse 5–13, 81377 Munich (Germany) E-mail: thomas.carell@cup.uni-muenchen.de
Dr. S. Michalakos, Prof. Dr. M. Biel Center for Integrated Protein Science (CiPS^M) at the Department of Pharmacy, Ludwig-Maximilians-University Butenandtstrasse 5–13, 81377 Munich (Germany)

[+] These authors contributed equally to this work.

[**] We thank the excellence cluster CiPS^M, SFB 646, and SFB 749 for generous support. M. Münzel is grateful to the Fonds of the Chemical Industry for a Kekulé fellowship. We thank V. Hammelmann for support during this project.



Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201002033>.

with the TBS-protected bromo-dT derivative **1**.^[9] S_N2 substitution with H₂O followed by protection with a TBS group furnished compound **2**. H₂¹⁸O was used for the labeled compound. Subsequent conversion of **2** into the corresponding dC derivative by activation as a sulfonate^[10] and subsequent replacement with NH₃ yielded ⁵-HOMe dC after final deprotection. For ⁵-Me dC we used TBS-protected dT and converted it into TBS-protected ⁵-Me dC by using the strategy described above. The isotope-labeled **6** was prepared from TBS-protected iodocytosine **5**^[11] by palladium-catalyzed methylation at C5. Final deprotection furnished [D₃]⁵-Me dC. These nucleosides were used to generate mass spectrometry calibration curves to enable precise LC-MS quantification (see the Supporting Information).^[12] The obtained areas of the ion currents at the exact masses of each nucleoside were plotted against different concentration ratios of natural to labeled nucleosides. In both cases linear plots were obtained with R² values of 0.999.

We next used the calibration curve equations to analyze the content of ⁵-HOMe dC and ⁵-Me dC in various mouse brain tissues (Figure 1a). To this end, we collected the hypothalamus, cortex, hippocampus, olfactory bulb, brainstem, cerebellum, and retina from four 90 day old mice. The tissues were homogenized and the DNA isolated by phenol/chloroform extraction (see the Supporting Information). From the obtained DNA, 4–10 μg were completely digested to the nucleosides in a two-step procedure by incubating first with nuclease S1 at 37 °C for 3 h. In the second incubation step phosphodiesterase I and antarctic phosphatase were used at 37 °C for 3 h to enable total digestion of the DNA. We then added the isotope-labeled compounds and analyzed the nucleoside mixture by HPLC and using a high-resolution mass spectrometer as detector (Thermo Finnigan LTQ Orbitrap XL). In all experiments ⁵-HOMe dC and ⁵-Me dC eluted with retention times of 12.3 min and 18.5 min, respectively. One signal for the natural (light) and one for the synthetic (heavy) compound was detected in each experiment. Quantification was performed by comparing the integrals of the ion current of the natural compound (determined amount) with their corresponding heavy atom labeled derivatives (known amounts) by using the calibration

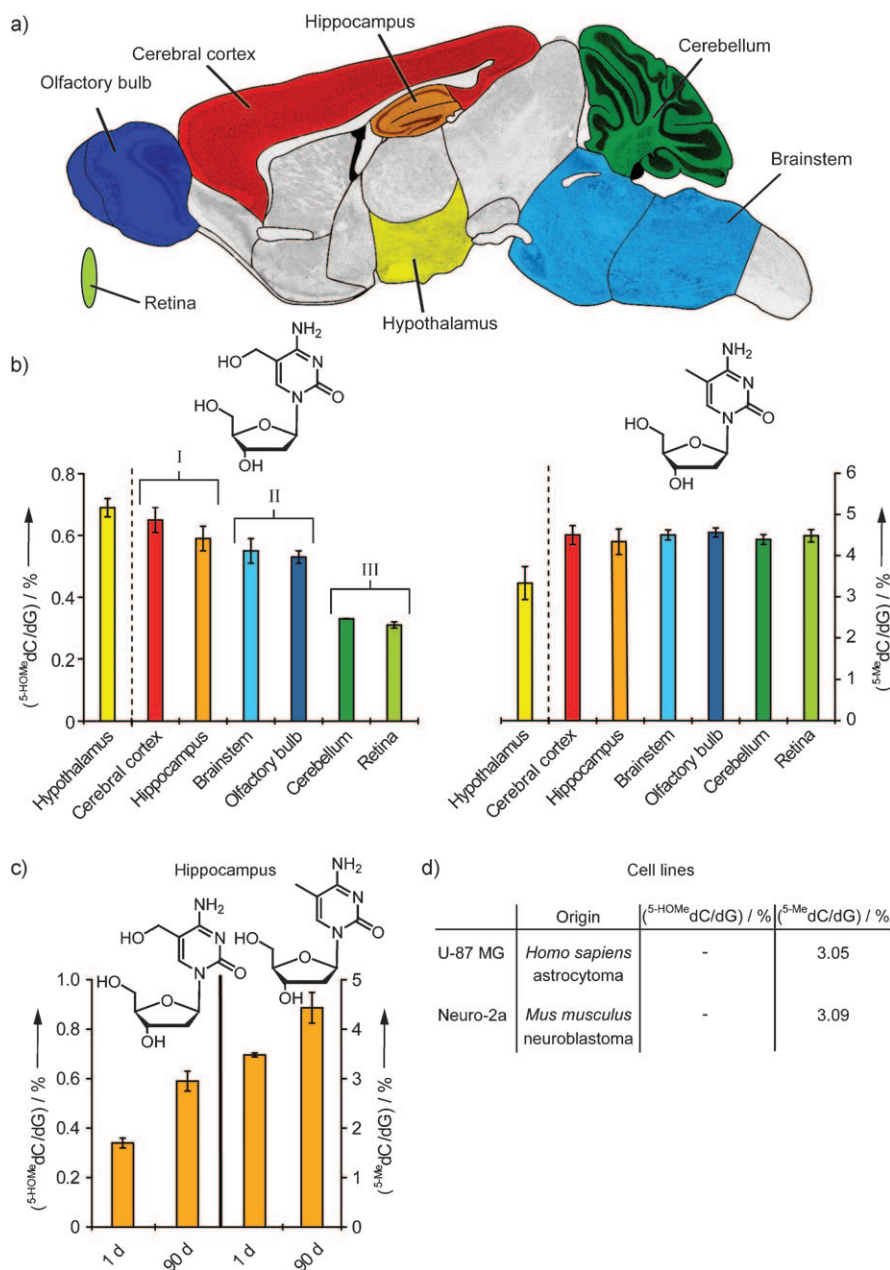


Figure 1. a) Sagittal section of a mouse brain. The brain areas highlighted in color were studied here. b) Ratio of ⁵-HOMe dC and ⁵-Me dC to dG in the different tissues in percent. dG was chosen as a reference, because it forms base pairs with dC, ⁵-HOMe dC, and ⁵-Me dC in DNA. I–III: see text. c) Ratio of ⁵-HOMe dC and ⁵-Me dC to dG in the hippocampus of 1 day and 90 day old mice in percent. d) Ratio of ⁵-HOMe dC and ⁵-Me dC to dG in a neuronal (Neuro-2a) and a glial (U-87 MG) cell line in percent.

curve equations. Repetition of the experiments allowed us to determine an average error for the quantification data of $\pm 5\%$ for ⁵-Me dC and ⁵-HOMe dC. All the data for each mouse are listed in the Supporting Information. Figure 1 shows the averaged value.

The first discovery (Figure 1b) is that the new base ⁵-HOMe dC is clearly detected in all brain tissues. Between 0.3 % and 0.7 % of all the dC nucleosides are hydroxymethylated at position 5, which clearly establishes ⁵-HOMe dC as a new post-replicatively formed nucleoside in the brain. The second

surprising result is that the amount of $^5\text{-HOME dC}$ varies significantly in the investigated tissues. The previous study^[5] found $^5\text{-HOME dC}$ predominantly in purkinje neurons, which are found in the cerebellum. We now see that the amount of the base is much larger in the cortex and hippocampus, where purkinje cells are not present. Analysis of brain cancer cell lines (Figure 1d) revealed that the new base is absent in neuroblastoma and glial cell lines. We found, however, and in agreement with the literature, reduced $^5\text{-Me dC}$ content (from a typical 4.5 % to 3 %) in these cancer cell lines.^[13]

By using our data we can roughly divide the mouse brain into three different areas (Figure 1b). Most $^5\text{-HOME dC}$ is found in the hippocampus and cortex (I), which are the brain areas that have higher cognitive functions. The brainstem and olfactory bulb form a second category, which possess intermediate $^5\text{-HOME dC}$ levels (II). The cerebellum and retina contain significantly less $^5\text{-HOME dC}$ (III). Surprisingly, initial data also show a relatively high level of $^5\text{-HOME dC}$ in the hypothalamus, which is part of the endocrine system that controls hormone-based processes. The distribution of $^5\text{-HOME dC}$ was confirmed by immunolocalization (see the Supporting Information).

The relative amounts of $^5\text{-HOME dC}$ and $^5\text{-Me dC}$ are also interesting. We are unable to correlate the amount of $^5\text{-HOME dC}$ and $^5\text{-Me dC}$ in these tissues, which one would expect if $^5\text{-HOME dC}$ is exclusively generated from $^5\text{-Me dC}$ as a precursor. The $^5\text{-Me dC}$ values are significantly lower in the hypothalamus and here we indeed detect high $^5\text{-HOME dC}$ values. In contrast, and in accord with the literature,^[14] we found that the $^5\text{-Me dC}$ values are stable at a typical value of around 4.5 % in all other tissues, while the $^5\text{-HOME dC}$ values vary significantly. This suggests that $^5\text{-HOME dC}$ has a function that is not correlated with the $^5\text{-Me dC}$ value.

Finally, we asked the question how the age of the animal influences the $^5\text{-Me dC}$ and $^5\text{-HOME dC}$ values. To this end, we analyzed the hippocampus tissue of one-day old mice (Figure 1c). We indeed detected a significantly lower level of both $^5\text{-Me dC}$ and $^5\text{-HOME dC}$. The $^5\text{-Me dC}$ value increases with age from $(3.5 \pm 0.1) \%$ to $(4.3 \pm 0.3) \%$. More significantly, the amount of $^5\text{-HOME dC}$ is raised in 90 day old mice by approximately 75 % from $(0.34 \pm 0.02) \%$ to $(0.59 \pm 0.04) \%$. To exclude that $^5\text{-HOME dC}$ accumulates because of oxidative stress in older animals we also analyzed the damaged base 8-oxodG, which is a typical oxidative stress marker. Here, however, a difference could not be detected, which shows that $^5\text{-HOME dC}$ in the brain does not result from an accumulation of oxidative DNA damage.^[15]

In summary, we have quantified the amount of $^5\text{-Me dC}$ and $^5\text{-HOME dC}$ in brain tissues with excellent accuracy and we confirmed $^5\text{-HOME dC}$ as a new post-replicatively formed

nucleoside. We were able to show that the base is widely distributed in the brain and that the nucleoside is particularly prominent in those brain tissues which are involved in higher cognitive functions. Finally, we obtained initial data that show that the $^5\text{-HOME dC}$ nucleoside is present in new born mice at reduced levels.

Received: April 6, 2010

Revised: May 3, 2010

Published online: June 25, 2010

Keywords: demethylation · DNA methylation · DNA modification · epigenetics · hydroxymethylcytosine

- [1] J. D. Watson, F. H. C. Crick, *Nature* **1953**, *171*, 737–738.
- [2] a) A. Bird, *Nature* **2007**, *447*, 396–398; b) P. A. Jones, D. Takai, *Science* **2001**, *293*, 1068–1070.
- [3] M. G. Goll, T. H. Bestor, *Annu. Rev. Biochem.* **2005**, *74*, 481–514.
- [4] a) A. P. Feinberg, *Nature* **2007**, *447*, 433–440; b) W. Reik, *Nature* **2007**, *447*, 425–432.
- [5] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930.
- [6] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930–935.
- [7] a) C. Loenarz, C. J. Schofield, *Chem. Biol.* **2009**, *16*, 580–583; b) Z. Liutkevičiūtė, G. Lukinavičius, V. Masevičius, D. Daujotytė, S. Klimašauskas, *Nat. Chem. Biol.* **2009**, *5*, 400–402; c) J.-K. Zhu, *Annu. Rev. Genet.* **2009**, *43*, 143–166.
- [8] For an alternative synthesis, see C. J. LaFrancois, J. Fujimoto, L. C. Sowers, *Chem. Res. Toxicol.* **1998**, *11*, 75–83.
- [9] a) R. K. Grover, S. J. K. Pond, Q. Cui, P. Subramaniam, D. A. Case, D. P. Millar, P. Wentworth, *Angew. Chem.* **2007**, *119*, 2897–2901; *Angew. Chem. Int. Ed.* **2007**, *46*, 2839–2843.
- [10] N. Bischofberger, M. D. Matteucci, *J. Am. Chem. Soc.* **1989**, *111*, 3041–3046.
- [11] a) C. H. Hwang, J. S. Park, J. H. Won, J. N. Kim, E. K. Ryu, *Arch. Pharmacol. Res.* **1992**, *15*, 69–72; b) M. Kuwahara, J.-I. Nagashima, M. Hasegawa, T. Tamura, R. Kitagata, K. Hanawa, S.-I. Hososhima, T. Kasamatsu, H. Ozaki, H. Sawai, *Nucleic Acids Res.* **2006**, *34*, 5383–5394.
- [12] a) T. Brückl, D. Globisch, M. Wagner, M. Müller, T. Carell, *Angew. Chem.* **2009**, *121*, 8074–8077; *Angew. Chem. Int. Ed.* **2009**, *48*, 7932–7934; b) E. P. Quinlivan, J. F. Gregory, *Nucleic Acids Res.* **2008**, *36*, e119; c) Y. Yang, D. Nikolic, S. M. Swanson, R. B. van Breemen, *Anal. Chem.* **2002**, *74*, 5376–5382.
- [13] M. Esteller, *N. Engl. J. Med.* **2008**, *358*, 1148–1159.
- [14] a) M. A. Gama-Sosa, R. M. Midgett, V. A. Slagel, S. Githens, K. C. Kuo, C. W. Gehrke, M. Ehrlich, *Biochim. Biophys. Acta Gene Struct. Expression* **1983**, *740*, 212–219; b) B. F. Vanyushin, N. G. Lopatina, C. K. Wise, F. R. Fullerton, L. A. Poirier, *Eur. J. Biochem.* **1998**, *256*, 518–527.
- [15] C. Bienvenu, J. R. Wagner, J. Cadet, *J. Am. Chem. Soc.* **1996**, *118*, 11406–11411.

4.5 “Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates”

D. Globisch, M. Münzel, M. Müller, S. Michalakis, **M. Wagner**, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS ONE* **2010**, 5, e15367.

Ziel dieser Arbeit war es, die zuvor von uns in Hirngeweben durchgeführte Quantifizierung von hmC und mC auf nicht-neuronale Gewebe bzw. weitere Organe auszudehnen. Die durch Oxidation von hmC gebildeten DNA-Basen fC und caC waren zu diesem Zeitpunkt noch nicht entdeckt worden. Es wurde jedoch schon spekuliert, dass fC und caC als mögliche Intermediate eines aktiven DNA-Demethylierungsprozesses auftreten könnten. Ein weiteres wichtiges Ziel dieses Projektes war es daher, in den analysierten Proben nach diesen beiden Basen und nach hmU, einem weiten möglichen Intermediat der aktiven DNA-Demethylierung, zu suchen.

In dieser Publikation haben wir gezeigt, dass hmC in allen untersuchten Organen der Maus vorkommt. Während das hmC-Level zwischen den Organen variiert und die mit Abstand höchsten Level im Gehirn gemessen wurden, wiesen alle Organe das gleiche mC-Level auf. Leider haben wir in dieser Arbeit keines der postulierten Intermediate der aktiven DNA-Demethylierung nachweisen können. Wie sich später herausgestellt hat, lag die Nachweisgrenze unserer Quantifizierungsmethode immer noch zu hoch. Sie hätte aber möglicherweise für einen Nachweis von fC in embryonalen Stammzellen ausgereicht, da der DNA-Formylierungsgrad dort höher ist als in den Geweben adulter Individuen.

Für dieses Projekt habe ich, zusammen mit Susanne Koch, für alle durchgeführten Messungen die DNA aus den Gewebeproben isoliert.

Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates

Daniel Globisch¹*, Martin Münzel¹, Markus Müller¹, Stylianos Michalakis², Mirko Wagner¹, Susanne Koch², Tobias Brückl¹, Martin Biel², Thomas Carell^{1*}

1 Center for Integrated Protein Science (CiPS^M) at the Department of Chemistry, Ludwig-Maximilians-University, Munich, Germany, **2** Center for Integrated Protein Science (CiPS^M) at the Department of Pharmacy, Ludwig-Maximilians-University, Munich, Germany

Abstract

5-Hydroxymethylcytosine (hmC) was recently detected as the sixth base in mammalian tissue at so far controversial levels. The function of the modified base is currently unknown, but it is certain that the base is generated from 5-methylcytosine (mC). This fuels the hypothesis that it represents an intermediate of an active demethylation process, which could involve further oxidation of the hydroxymethyl group to a formyl or carboxyl group followed by either deformylation or decarboxylation. Here, we use an ultra-sensitive and accurate isotope based LC-MS method to precisely determine the levels of hmC in various mouse tissues and we searched for 5-formylcytosine (fC), 5-carboxylcytosine (caC), and 5-hydroxymethyluracil (hmU) as putative active demethylation intermediates. Our data suggest that an active oxidative mC demethylation pathway is unlikely to occur. Additionally, we show using HPLC-MS analysis and immunohistochemistry that hmC is present in all tissues and cell types with highest concentrations in neuronal cells of the CNS.

Citation: Globisch D, Münzel M, Müller M, Michalakis S, Wagner M, et al. (2010) Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates. PLoS ONE 5(12): e15367. doi:10.1371/journal.pone.0015367

Editor: Anna Kristina Croft, University of Wales Bangor, United Kingdom

Received: September 17, 2010; **Accepted:** November 11, 2010; **Published:** December 23, 2010

Copyright: © 2010 Globisch et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Funding was received from the following sources: the excellence cluster CiPSM, the SFBs 646 and 749, the Fond der Chemischen Industrie and Bayer Schering Pharma AG. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Funding was received from Bayer-Schering Pharma AG, which does not alter the authors' adherence to the policies of PLoS ONE.

* E-mail: Thomas.carell@cup.uni-muenchen.de

† These authors contributed equally to this work.

Introduction

In 2009 it was discovered that the genetic material contains aside from the nucleobases A, C, G, T, and 5-methylcytosine (mC) the additional sixth base 5-hydroxymethylcytosine (hmC) [1,2]. Whereas it is well established that mC is a crucial epigenetic marker [3], the function of the sixth base hmC is currently unknown, but it is speculated that it is involved in epigenetic regulation events or in active demethylation processes [4,5,6,7,8]. It was discovered that the base is generated from mC as precursor via oxidation (hydroxylation) by members of the TET enzyme family [2,9]. These TET enzymes are Fe(II)-dependent α -oxoglutarate dioxygenases, which utilize molecular oxygen to hydroxylate alkyl groups to hydroxymethyl functionalities *via* a radical based mechanism [10]. It is currently hypothesized that hmC could be a base involved in epigenetic modulation of gene activity. The fact that hmC was also discovered in embryonic stem cells (ES cells) and seems to play an important role in ES cells self-renewal supports this hypothesis [9].

The base was originally detected in several tissues with the highest levels in mammalian brain. In brain tissues, between 0.4% and 0.7% of all dC bases were found to be converted into hmC [1]. These values are in good agreement with a recent, more detailed LC-MS study developed in our laboratory [11]. The amount of hmC in other tissues was initially found to be small with levels up to 0.1% but more precise quantification could not be obtained due to the sensitivity limitations associated with the early detection method. In an alternative attempt to achieve better

information about the hmC content, an enzymatic approach was taken in which UDP-[³H]glucose was used as a marker, which was transferred onto hmC with the help of a specific glucosyltransferase. This assay provided hmC values far higher than those reported earlier [12].

hmC could in principle directly modulate the binding of proteins to DNA also during epigenetic regulation processes. For example it was shown that methyl-CpG binding protein 2 (MeCP2) no longer binds to the corresponding sequences when mCs were converted to hmCs in CpG sequences [13]. Alternatively, it could function as an intermediate in active oxidative demethylation (**Figure 1**) [14]. Oxidation of the hydroxymethyl group to a formyl group would yield 5-formylcytosine (fC) which could expel formic acid and react to C. Another possibility would be further oxidation of hmC to 5-carboxylcytosine (caC) which possesses a carboxyl group and would enable quick decarboxylation to regenerate dC (**Figure 1**). Nature's most proficient enzyme orotate decarboxylase catalyzes a similar reaction in which orotate is decarboxylated to uridine [15]. Similar oxidation and decarboxylation reactions are known for thymine in the pyrimidine salvage pathway of certain eukaryotes [16,17,18]. Alternatively, hmC or its deamination product hmU could be substrates of specialized DNA glycosylases [6,19,20,21]. Recent observations that epigenetic reprogramming is associated with the activation of base excision repair (BER) pathways support this hypothesis [22,23]. During preparation of this manuscript a review was published, which independently postulated these two putative pathways for active DNA demethylation [24].

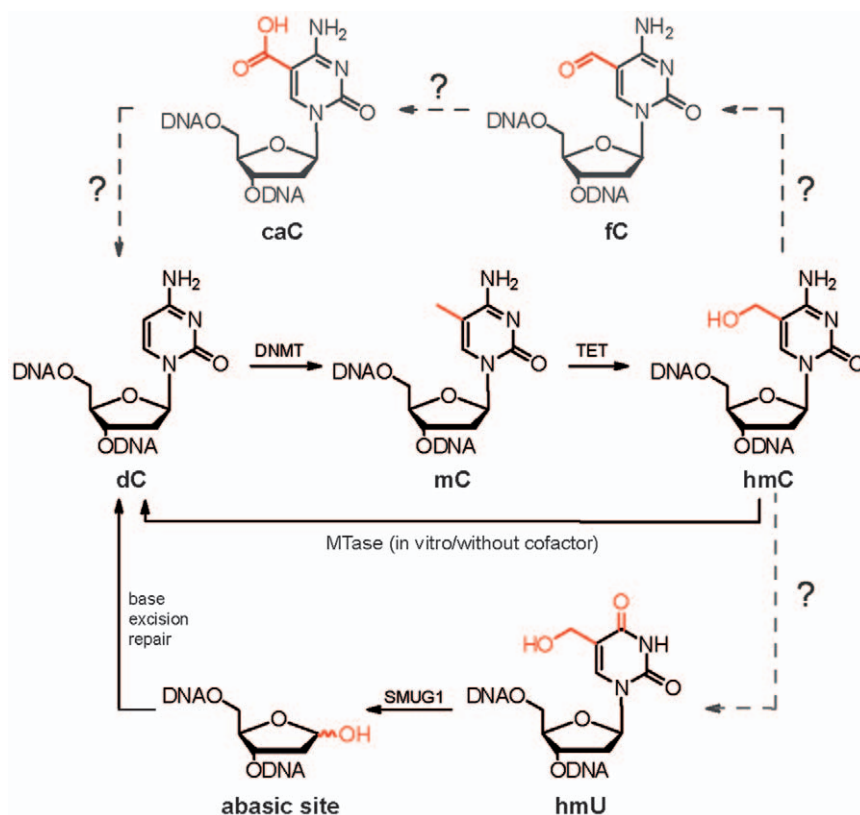


Figure 1. Putative demethylation pathways. Depiction of the known cytosine modifications mC and hmC and of the putative oxidative “demethylation” intermediates fC and caC. The base excision repair (BER) pathway is a second possible demethylation pathway via the intermediate hmU.

doi:10.1371/journal.pone.0015367.g001

Results

In order to test the idea that hmC is an intermediate of such an oxidative demethylation pathway and to shine light on the distribution of hmC in mammalian tissues we applied an isotope based HPLC-MS method to quantify hmC and to detect the presence of fC, caC, and hmU in various tissues. Mass spectrometry is a highly sensitive detection method that is, however, not quantitative. (In certain cases mass spectrometry was used for quantification of nucleosides [25]). Using stable isotope-labeled reference compounds, which have the same chromatographic and ionization properties but different molecular weights, turns MS into a quantitative method [11,26,27,28]. By comparing the integrals of the individual mass signals of the natural compound (amount to be determined) and labeled compound (known amount) and subsequent application of calibration curves (**Figure S1**) very precise quantification of the natural compound is achieved.

Synthesis of modified cytosine derivatives

The syntheses of the putative intermediates fC, caC, and hmU are depicted in **Figure 2**. Starting point was dC, which was first iodinated at position 5 and subsequently TBS protected to yield **1** [29]. A Pd-catalyzed carbonylation with Bu_3SnH gave formylated compound **2**. This reaction proceeded in excellent yield even in the presence of the unprotected amine at position 4. Deprotection of **2** with HF in pyridine generated fC. hmC was easily obtained after reduction of **2** and subsequent deprotection of the TBS groups. Here, addition of a Lewis acid [30] was essential, because simple reduction using NaBH_4 resulted mainly in decomposition of the starting

material, presumably because the hydride added to the position 6 of intermediate **2**. For the synthesis of the isotope-labeled standard $[\text{D}_2]$ -hmC for mass spectrometric analyses Bu_3SnD and NaBD_4 were used in the formylation and reduction reactions, respectively (for details refer to the **Text S1** and for a protective group free synthesis see **Figure S2**). For the synthesis of caC we converted iodine **1** to a TMS-ethanol-ester in a Pd-catalyzed reaction to yield compound **3** and subsequently cleaved the protective groups. Fully protected **4** was synthesized according to a literature procedure and converted to hmU with HF·pyridine [11].

HPLC-MS Quantification of hmC and mC in Mammalian Tissue

Using the LC-MS method, we quantified the levels of hmC and mC in parallel using the isotope-labeled derivative $[\text{D}_2]$ -hmC. This reference compound is labeled to a very high extent (>99%), which further enhances quantification accuracy of small amounts of hmC allowing us to determine the content of hmC in any tissue very precisely. For the study we isolated DNA from a variety of different mouse tissues [11]. The isotope-labeled compound $[\text{D}_2]$ -hmC and the obtained mixture was analyzed by HPLC-MS using a sensitive MS-detector (*Thermo Scientific LTQ Orbitrap XL*). The workflow of our method is illustrated in **Figure 3**. The distribution of hmC and mC in different mouse tissues is depicted in **Figure 4A and 4B**. We measured uniform amounts of mC that represent $4.30 \pm 0.22\%$ of dG in all tissues in agreement with previous reports [11,30,31]. The only exception is nasal epithelia with a slightly lower value. To our surprise we found that hmC is present in all investigated tissues at significant quantities. More

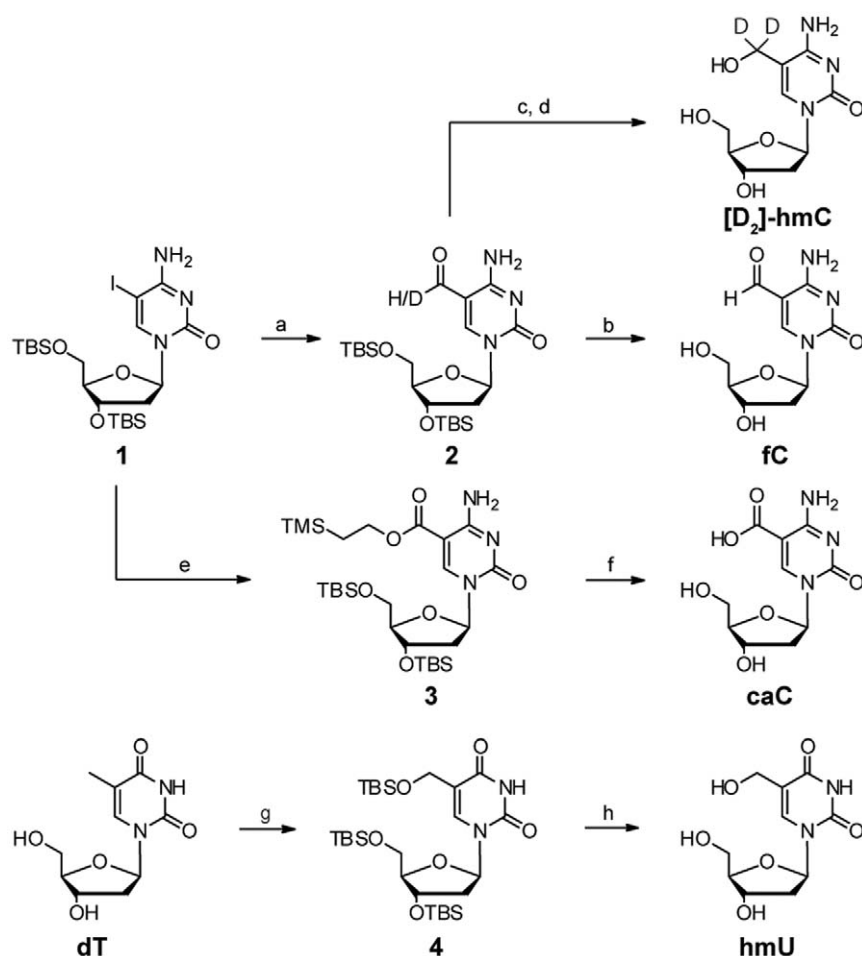


Figure 2. Synthesis of $[D_2]$ -hmC and the putative intermediates fC, caC, and hmU. a) CO, PPh_3 , $Pd_2(dba)_3 \cdot CHCl_3$, Bu_3SnH , 97%, with Bu_3SnD 49%, b) HF pyridine, 75%, c) $NaBD_4$, $CeCl_3 \cdot 7H_2O$, 21%, d) TBAF, 51%, e) CO, TMS-Et-OH, DIPEA, $Pd_2Cl_2(MeCN)_2$, 52%, f) TBAF, 69%, g) Reference [11] h) HF-pyridine, 70%. All reactions could also be carried out in a protective group free manner but resulted in reduced yields and tedious workups (see Figure S2 for details).

doi:10.1371/journal.pone.0015367.g002

important is the discovery that the hmC values are deviating strongly in contrast to the stable amounts of mC. Three different classes of tissue are clearly distinguishable. The highest levels (0.3%–0.7%) of hmC are detected in DNA isolated from the central nervous system (CNS). The spinal cord, which is not a direct part of the brain, also belongs to hmC rich tissues with a

value of 0.47%. These data strengthen the observation that neuronal tissues contain the highest levels of hmC. DNA extracted from kidney, nasal epithelium, bladder, heart, skeletal muscle, and lung has medium hmC values from 0.15%–0.17%. DNA from liver, spleen, and the endocrine glands (testes and pituitary gland) possess the lowest amounts of hmC with levels ranging from

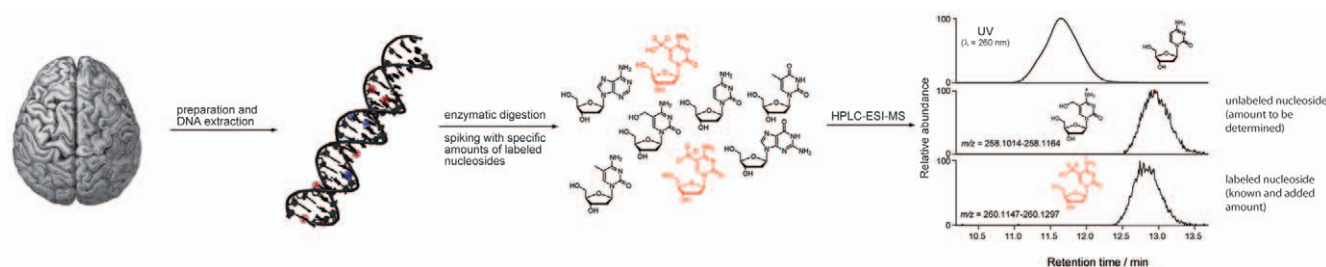


Figure 3. Workflow of the HPLC-MS quantification method. DNA is extracted from any kind of tissue and subsequently enzymatically digested to the nucleosides. Subsequently, a known amount of the stable isotope-labeled standard nucleoside is added. In the HPLC-MS analysis one signal for the natural (light) and one for the synthetic (heavy) compound is detected in each experiment. Quantification is performed by comparing the integrals of the specific high resolution ion current of the natural compound (amount to be determined) with their corresponding heavy atom labeled derivative (known amount).

doi:10.1371/journal.pone.0015367.g003

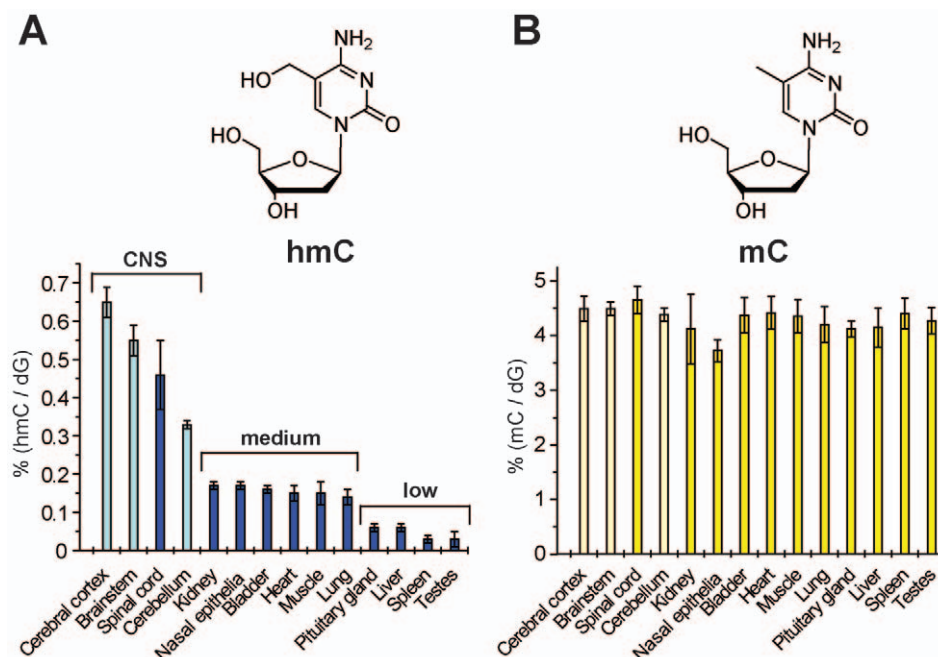


Figure 4. Quantification of hmC and mC in mouse tissue. A) Values of hmC in % of dG. B) Values of mC in % of dG. A)+B) Data represent values for each tissue of at least two mice with standard deviation (SD) (See **Table S1** for details). Light-colored bars represent data from our earlier study [11].

doi:10.1371/journal.pone.0015367.g004

0.03%–0.06%. This last class of tissues contains up to five times less hmC compared to tissues with medium levels and up to 20 times less than cerebral cortex as part of CNS. Interestingly, the pituitary gland, which is located in the brain, has a low hmC value of only 0.06%, supporting the hypothesis that high hmC content is related to neuronal function, rather than mere localization in the brain. In contrast to mC, the hmC amounts are tissue specific.

Immunohistology experiments

In order to further validate our quantification data and to determine the exact location of hmC in tissues more precisely we performed immunostaining experiments with a commercially available hmC-specific antibody [9,11]. The results are depicted in **Figure 5**. The pictures show, that the sixth base is clearly located in the cell nuclei as expected. Upon addition of a hmC containing capturing oligonucleotide the signal obtained from the nuclei is reduced showing that the antibody is specific for hmC. It is evident that the highest amounts of hmC are present in the nuclei of the hippocampus and that kidney is stained with a clearly higher intensity than liver, which supports our HPLC-MS results (See **Figure S3** for further immunohistology experiments). Furthermore, virtually all cells contain hmC. Interestingly, whereas hmC is equally distributed in liver and kidney, its location in the hippocampus is very diverse. The highest levels are detected in the fully differentiated neurons of the dentate gyrus. Cells located in the subgranular zone between dentate gyrus and hilus show clearly reduced staining in line with reduced hmC levels (**Figure 6**). This area contains the neural progenitors known to generate new neurons in the hippocampus of adult mice [32,33].

Putative demethylation pathways

To seek evidence if hmC is part of an oxidative demethylation pathway we investigated if the DNA samples contain fC and caC as putative downstream intermediates. Using the LC-MS method

we first determined the detection limit to be in the low picomolar range, even with the background of canonical nucleosides in digested samples (**Figure 7B**). This shows that even traces of these compounds at levels 70–350 times lower than the base hmC itself could unambiguously be detected. The HPLC-chromatogram of caC, hmC, hmU, mC, and fC is shown in **Figure 7A**. In addition we maximized the amount of digested DNA applicable for HPLC-MS analysis. For these experiments, we typically used 70–160 μ g of DNA isolated from the tissues. The DNA was again enzymatically digested and analyzed. For comparison, the previous experiments were performed using ten times less DNA. Despite this, neither fC nor caC could be detected. The results show that if present, fC does not reach levels above $7 \cdot 10^{-4}\%$ of all nucleosides and 0.3% of hmC and caC does not reach levels above $3.5 \cdot 10^{-3}\%$ of all nucleosides and 1.4% of hmC. Our results show that either an oxidative active demethylation pathway does not exist, that intermediates are short lived or that they are not released from the enzymatic complex.

Another potential active demethylation pathway could involve the base excision repair (BER) pathway of cells (**Figure 1**) either directly or via hmU [21]. Thus we investigated the presence of hmU, which could be formed by deamination of hmC by an analog of activation induced cytidine deaminase (AID) [6]. hmU is known to be a substrate for the glycosylase SMUG1 [19,20]. It can be speculated that deamination of hmC to give hmU could be a signal for repair driven removal of mC via hmC. Again our HPLC-MS method did not provide any evidence for the presence of this compound in the digested DNA making such a mechanism equally unlikely (**Figure 7B**).

Discussion

The role of mC in epigenetic regulatory processes is today quite well understood. The introduction and heredity of methylation patterns by DNA methyltransferases is an established fact.

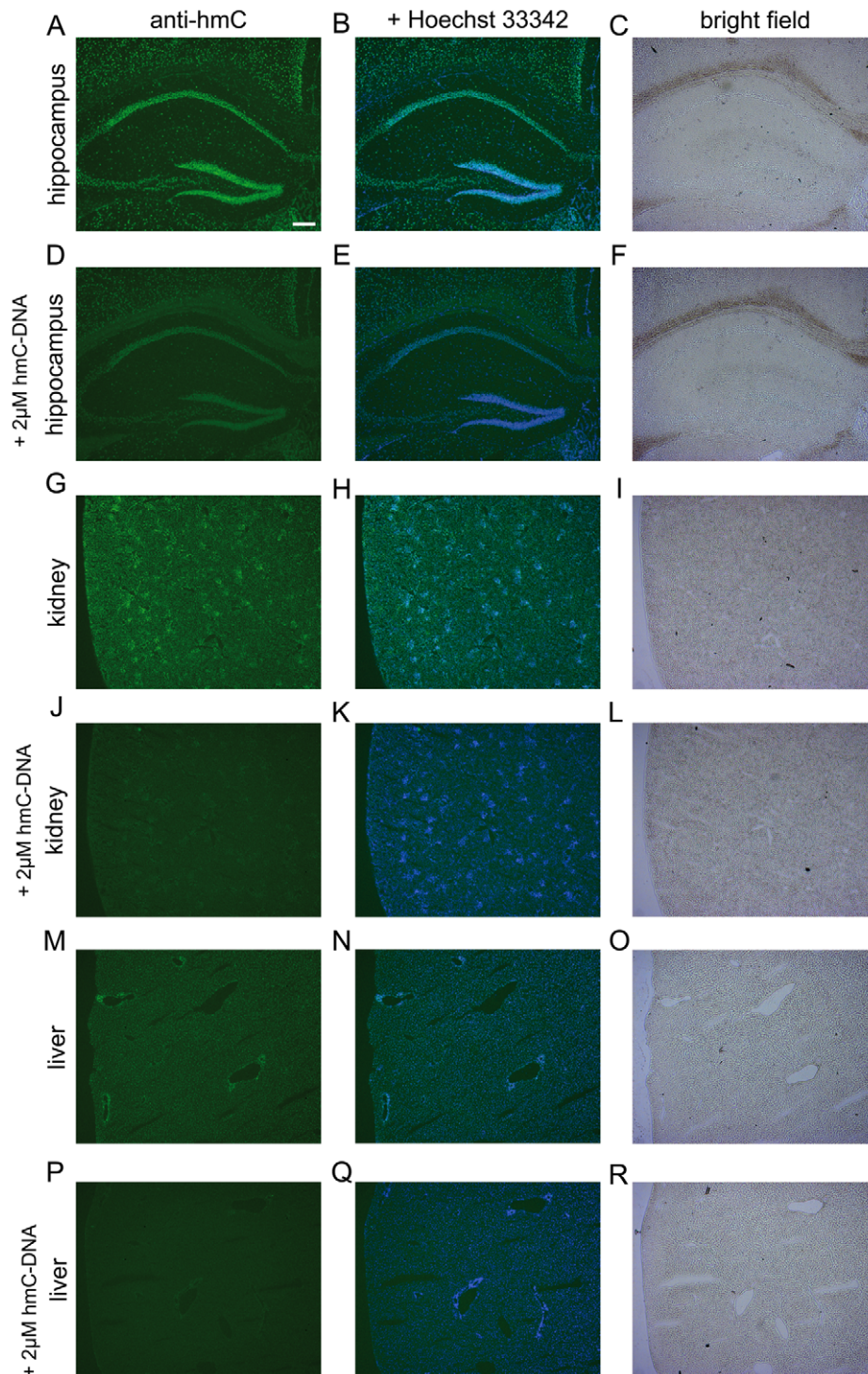


Figure 5. Immunolocalization of hmC in mouse hippocampus, kidney, and liver. Scale bar: 200 μ M. Left column: mouse tissues stained with anti-hmC (green). Middle column: mouse tissues stained with anti-hmC (green) and Hoechst 33342 (blue) for nuclear staining. Right column: Bright field pictures of corresponding tissue. In every second row 2 μ M hmC-DNA were added to compete the anti-hmC staining signal out. doi:10.1371/journal.pone.0015367.g005

However, the removal of these epigenetic marks from the genetic code is still controversially discussed [24]. Thus it is not surprising that the discovery of hmC and the modifying TET enzymes in 2009 caused significant excitement. To elucidate the function of hmC in mammalian tissue, the knowledge of its distribution could provide valuable information.

How is hmC distributed in mammalian tissues?

Our quantification results in different tissues show that hmC is present in every tissue investigated. However, our obtained values are in disagreement with the recently determined higher levels measured using an enzymatic method [12] but they are in good agreement with those levels reported initially based on radioactive

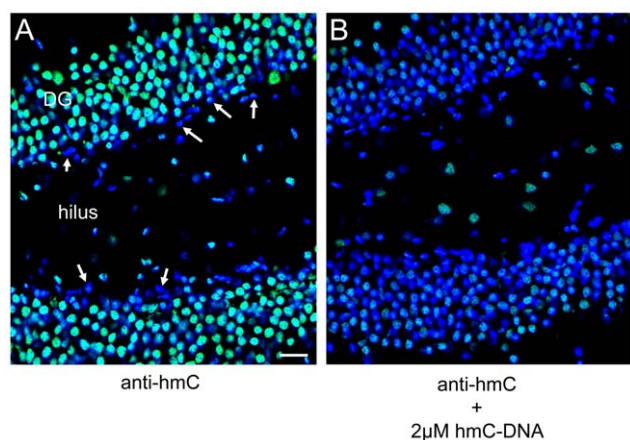


Figure 6. Immunolocalization of hmC in mouse hippocampus. High magnification images of hmC immunoreactivity in the dentate gyrus (DG) and the hilus of mouse hippocampus. A) Signal for anti-hmC (green) and Hoechst 33342 nuclear dye (blue) B) Competition of anti-hmC with 2 μ M hmC-DNA. The scale bar marks 20 μ m. doi:10.1371/journal.pone.0015367.g006

labeled gel chromatography even though the estimated levels were at that time close to the detection limit of the method [1]. By employing our HPLC-MS method, that uses an internal standard and calibration curves, we solve this scientific controversy and show that hmC is present in cells in levels between 0.03% and 0.69% of dG.

Our immunohistology data support the idea that oxidation of mC to hmC is a process occurring in all cells. Interestingly, we found indication in the dentate gyrus of the hippocampus that the hmC levels are significantly lower in stem cell rich areas.

Is hmC an intermediate in active cytosine demethylation?

It was speculated that hmC could represent an intermediate of an oxidative mC removal pathway. We proposed that hmC can be further oxidized to fC and caC similar to the pyrimidine salvage pathway or that it could be deaminated to hmU to be further excised by BER enzymes. Based on our data, however, we can exclude substantial oxidation of hmC to fC or caC or deamination of hmC to result in hmU. This does not disprove that these pathways do exist since we cannot exclude that the application of more sensitive methods, such as HPLC-MS/MS may lead to the discovery of these modifications in the future. In addition, if hmC or hmU were excised directly from DNA, it should be possible to detect these degradation products in urine [34]. But the data show that these unavoidable intermediates do not accumulate to any significant level, making it unlikely that such reactions occur at large scale or that they are extremely short lived intermediates.

Materials and Methods

General methods

All non-aqueous reactions were performed using flame- or oven-dried glassware under an atmosphere of dry nitrogen. Commercial reagents from Sigma-Aldrich or Acros were used as received unless otherwise noted. Non-aqueous reagents were transferred under nitrogen with a syringe or cannula. Solutions were concentrated *in vacuo* on a Heidolph rotary evaporator. Chromatographic purification of products was accomplished using flash column chromatography on Merck Geduran Si 60 (40–63 μ m) silica gel (normal phase), Fluka silica gel 100 C₁₈-Reversed phase (15–35 μ m), or preparative HPLC on a Merck-Hitachi system (L-7150 pump, L-7420 detector) equipped with a Macherey Nagel VP 250/32 Nucleosil 100-7 C₁₈ column, 15.0 mL/min flow rate, as indicated. Thin layer chromatography (TLC) was performed on Merck 60 (silica gel F254) plates. Visualization of the developed

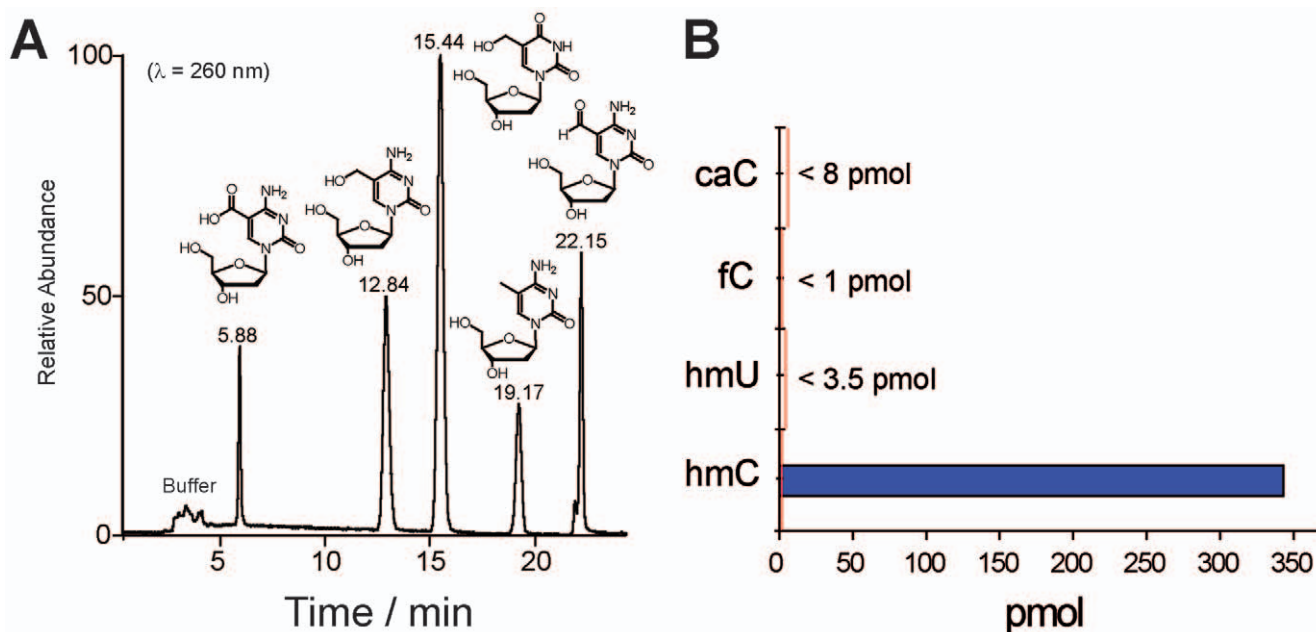


Figure 7. Detection of potential demethylation intermediates caC, hmU, and fC. A) HPLC-chromatogram of the synthesized cytosine and uracil modifications caC, hmC, hmU, mC, and fC as 2'-deoxynucleosides showing excellent separation of the compounds. B) Detected values of the potential intermediates as example in olfactory bulb. The red line indicates the detection limits of the modified nucleosides in enzymatically digested DNA samples. doi:10.1371/journal.pone.0015367.g007

chromatogram was performed using fluorescence quenching or anisaldehyde staining. ^1H and ^{13}C NMR spectra were recorded in deuterated solvents on *Bruker ARX 300*, *Varian VXR400S*, *Varian Inova 400* and *Bruker AMX 600* spectrometers and calibrated to the residual solvent peak. Multiplicities are abbreviated as follows: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet. ^{13}C -NMR signals of carbons carrying deuterium atoms have been determined by C-H correlation spectra (HMBC). ESI spectra and high-resolution ESI spectra were obtained on the mass spectrometer *Thermo Finnigan LTQ FT-ICR*. Acetonitrile (HPLC gradient grade) for HPLC-ESI-MS analysis was purchased from VWR. HCOOH (p.a. for mass spectrometry) was purchased from Fluka. Mouse tissues from three male mice (C57BL/6N) were frozen in liquid nitrogen right after sacrifice. Depending on the tissue type each mouse supplied enough DNA for up to 4 measurements.

DNA extraction from mouse tissue

DNA isolation was performed on the basis of the QIAamp DNA Mini Kit. Instead of column purification phenol extraction was performed as outlined in the next paragraph. The RNA digest was executed twice. All other steps were performed as described by the manufacturer. For samples with more than 25 mg weight the quantities of the reagents were increased accordingly. Tissue samples were homogenized with PBS and a stainless steel bead in a TissueLyser (Qiagen, 30 Hz, 2 min). ATL and proteinase K were added and the solution was incubated. DNase-free RNase A (4 μL , 100 mg/mL) was added. After mixing the sample was incubated and shaken (600 rpm) at rt for 5 min. A second portion of DNase-free RNase A (4 μL , 100 mg/mL) was added and the mixture was again incubated and shaken (600 rpm) at rt for 5 min. The tube was centrifuged briefly and buffer AL was added. The sample was mixed and incubated. Following this step the sample was no longer processed on the basis of the QIAamp DNA Mini Kit.

The sample was distributed equally to two 2 mL reaction tubes, if necessary. A 1/1 mixture of Roti®Phenol/chloroform (1 Vol.) was added and the tube was shaken vigorously at rt for 5 min. The tube was centrifuged (12100 g, 15 min) and the aq. layer was collected. This procedure was repeated once. To the obtained aq. layer chloroform (1 Vol.) was added and the tube was shaken at rt for 1 min. After centrifugation (12100 g, 5 min) the aq. layer was collected. During collection of the aq. layers special care was taken to include the interphase. The sample was distributed equally to two 2 mL reaction tubes, if necessary. Ethanol (3 Vol.) was added. The sample was left to stand at rt for approximately 2 h. After precipitation of the DNA the tube was centrifuged (12100 g, 30 min). The supernatant was discarded and the pellet was dried. Subsequently, it was dissolved in water (100–400 μL). The solution was centrifuged (12100 g, 30 min) and the supernatant was collected.

Enzymatic digestion

For the enzymatic digestion DNA mixtures (4 to 10 μg in a final volume of 100 μL H_2O) were heated to 100°C for 5 min to denature the DNA and rapidly cooled on ice. Buffer A (10 μL , 300 mM ammonium acetate, 100 mM CaCl_2 , 1 mM ZnSO_4 , pH 5.7) and nuclease S1 (80 units, *Aspergillus oryzae*) were added to the mixture and incubated for 3 h at 37°C. Addition of buffer B (12 μL , 500 mM Tris-HCl, 1 mM EDTA), antarctic phosphatase (10 units), snake venom phosphodiesterase I (0.2 units, *Crotalus adamanteus venom*) and incubation for further 3 h at 37°C completed the digestion. Labeled nucleosides $[\text{D}_2]\text{hmC}$ and $[\text{D}_3]\text{mC}$ were added, followed by centrifugation of the sample (12100 g, 15 min). The supernatant was removed, the volume reduced to 100 μL and measured with HPLC-ESI-MS. Each sample was analyzed at least

in duplicate with independent concentrations of the two labeled nucleosides. The concentrations of standard solutions were chosen to be in the expected range of the sample nucleoside concentration.

HPLC-ESI-MS

The samples (100 μL injection volume) were analyzed by HPLC-ESI-MS on a *Thermo Finnigan LTQ Orbitrap XL* and chromatographed by a *Dionex Ultimate 3000 HPLC* system with a flow of 0.15 mL/min over an Uptisphere120-3HDO column from *Interchim*. The column temperature was maintained at 30°C. Eluting buffers were buffer C (2 mM HCOONH_4 in H_2O (pH 5.5)) and buffer D (2 mM HCOONH_4 in $\text{H}_2\text{O}/\text{MeCN}$ 20/80 (pH 5.5)). The gradient was 0 \rightarrow 12 min; 0% \rightarrow 3% buffer D; 12 \rightarrow 60 min; 3% \rightarrow 60% buffer D; 60 \rightarrow 62 min; 60% \rightarrow 100% buffer D; 62 \rightarrow 70 min; 100% buffer D; 70 \rightarrow 85 min; 100 \rightarrow 0% buffer D; 85 \rightarrow 95 min; 0% buffer D. The elution was monitored at 260 nm (*Dionex Ultimate 3000 Diode Array Detector*). The chromatographic eluent was directly injected into the ion source without prior splitting. Ions were scanned by use of a positive polarity mode over a full-scan range of m/z 200–1000 with a resolution of 30,000. Parameters of the mass spectrometer were tuned with a freshly mixed solution of adenosine (5 μM) in buffer C. The parameters used in this section were sheath gas flow rate, 16 arb; auxiliary gas flow rate, 11 arb; sweep gas flow rate, 4 arb; spray voltage, 5.0 kV; capillary temperature, 200°C; capillary voltage, 12 V, tube lens 60 V.

Mass calibration curves

Mass calibration curves of the labeled and corresponding unlabeled synthesized nucleosides were obtained at five different concentration ratios. For each concentration an average value of three independent measurements was determined (**Figure S1**). Each labeled nucleoside solution was mixed with three solutions with different concentrations of the corresponding unlabeled nucleosides. The areas of labeled and unlabeled nucleosides of the LC-MS measurements were determined using the *Qualbrowser* program by extraction of the accurate mass with a mass filter (**Table S1**) from the total ion current (TIC). The linear fits of the determined area ratios over the amount ratios gave R^2 -values of minimum 0.9987 for mC and 0.9997 for hmC. The linear equations were used for calculation of the exact nucleoside contents in genomic DNA samples. Synthetic labeled nucleosides were added to the enzymatically digested nucleoside mixture and the areas of labeled and unlabeled nucleosides were determined as described above. The amount of each nucleoside was calculated from the obtained area ratios and the linear fit equations of the calibration curves. Determination of the precise amounts of mC and hmC by the described isotope-labeled mass spectrometric method and determination of the precise amount of dG by integration of the UV/vis signal of the HPLC chromatogram enabled calculation of the exact percentage of mC and hmC to dG [11]. We have chosen dG as a standard, because it forms base pairs with all three cytosine derivatives dC, mC and hmC.

Immunohistochemistry

Coronal cryosections (12 μm) from 12 week old C57-BL6/N mice were rehydrated in phosphate buffered saline (PBS), fixed (10 min, 4% paraformaldehyde in PBS, pH 7.4), treated with 2N HCl in PBS (20 min) and incubated for 16 hours (4°C) with primary antibodies in 5% chemiblocker (Millipore, Germany) and 0.3% Triton X-100 in PBS. The primary antibody used was: rabbit anti-5-hydroxymethylcytosine (hmC, 1:500, Active Motif, Belgium). For secondary detection we used goat Alexa488 anti-

rabbit (1:800, Cell Signaling Technologies, Germany). Cell nuclei were counterstained with Hoechst 33342 (5 µg/mL, Sigma, Germany) and sections were mounted with aqueous mounting medium (PermaFluor, Beckman-Coulter, USA). Tissues were analyzed using a Zeiss AxioScope epifluorescence microscope equipped with a HBO 100 mercury arc lamp, appropriate filters equipped with an MRc ccd camera (Zeiss, Germany). Laser scanning confocal micrographs were collected using a LSM 510 meta microscope (Carl Zeiss, Germany).

Supporting Information

Figure S1 Mass calibration curves of the nucleosides mC and hmC. Linear fits of five data points represent perfect linearity with R^2 -values of 0.9987 for mC and 0.9997 for hmC. (TIF)

Figure S2 Protective group free synthesis of putative cytosine derivatives. a) I_2 , mCPBA, 63% b) CO, Bu_3SnH , $Pd_2(dba)_3$, 35% c) $NaBH_4$, $CeCl_3 \cdot 7H_2O$, 53% d) TMS-Et-OH, CO, DIPEA, $PdCl_2(MeCN)_2$, 22% e) HF·pyr, 11%. The putative oxidative catabolites of hmC could also be synthesized in a protective group free manner. However, these syntheses suffer from lower yields and the workup procedures get considerably more tedious than with the TBS-protected compounds. dC can be converted to **5** by oxidative iodination using iodine and mCPBA. Subsequent palladium catalyzed formylation with CO and Bu_3SnH in DMF as solvent gave fC in 34% yield. Subsequent Luche reduction furnished hmC. **5** could be converted to **6** in a

carbonylating coupling of the iodine with TMS-ethanol. Final cleavage of the ester with F^- yields ethene, TMSF and CaC_2 . (TIF)

Figure S3 Immunolocalization of hmC in mouse kidney, liver, and heart. Scale bar: 20 µm. Mouse tissues stained with anti-hmC (green), Hoechst 33342 nuclear staining is shown in blue. A+G+M) anti-hmC staining. B+H+N) anti-hmC (green) and Hoechst 33342 (blue). D+J+P) The anti-hmC staining signal was competed out by 2µM hmC-DNA. E+K+Q) The anti-hmC staining signal was competed out by 2µM hmC-DNA. C+F+I+L+O+R) Bright field pictures of the used tissues. gl: glomerulus, s: sinusoid. (TIF)

Table S1 Nucleoside percentages of mC and hmC to dG in different mouse tissues (Figure 2 in main text). Each mouse tissue is listed with the determined values, the standard deviation (SD) and the relative standard deviation (RSD). The average mouse values are listed in the last column. For muscle, the DNA of mouse 1+2 was mixed due to a low amount of DNA. The applied mass ranges of analyzed nucleosides are shown as well. (PDF)

Text S1 Supporting text. (PDF)

Author Contributions

Conceived and designed the experiments: TC MB M. Müller. Performed the experiments: DG M. Münzel M. Müller SM MW SK. Analyzed the data: TC DG M. Münzel SM M. Müller. Contributed reagents/materials/analysis tools: M. Münzel TB SK. Wrote the paper: TC M. Münzel DG.

References

- Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain. *Science* 324: 929–930.
- Tahiliani M, Koh KP, Shen YH, Pastor WA, Bandukwala H, et al. (2009) Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* 324: 930–935.
- Law JA, Jacobsen SE (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat Rev Genet* 11: 204–220.
- Liutkeviciute Z, Lukinavicius G, Masevicius V, Daujotyte D, Klimasauskas S (2009) Cytosine-5-methyltransferases add aldehydes to DNA. *Nat Chem Biol* 5: 400–402.
- Miller CA, Sweatt JD (2007) Covalent modification of DNA regulates memory formation. *Neuron* 53: 857–869.
- Popp C, Dean W, Feng SH, Cokus SJ, Andrews S, et al. (2010) Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* 463: 1101–1126.
- Reik W, Dean W, Walter J (2001) Epigenetic reprogramming in mammalian development. *Science* 293: 1089–1093.
- Zhu JK (2009) Active DNA demethylation mediated by DNA glycosylases. *Annu Rev Genet* 43: 143–166.
- Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, et al. (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466: 1129–1133.
- Loenarz C, Schofield CJ (2009) Oxygenase catalyzed 5-methylcytosine hydroxylation. *Chem Biol* 16: 580–583.
- Münzel M, Globisch D, Brückl T, Wagner M, Welzmler V, et al. (2010) Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew Chem Int Ed* 49: 5375–5377.
- Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H (2010) Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res* 38: e181.
- Valinluck V, Tsai HH, Rogstad DK, Burdzy A, Bird A, et al. (2004) Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res* 32: 4100–4108.
- Privat E, Sowers LC (1996) Photochemical deamination and demethylation of 5-methylcytosine. *Chem Res Toxicol* 9: 745–750.
- Houk KN, Lee JK, Tantillo DJ, Bahmanyar S, Hietbrink BN (2001) Crystal structures of orotidine monophosphate decarboxylase: Does the structure reveal the mechanism of nature's most proficient enzyme? *ChemBiochem* 2: 113–118.
- Simmons JM, Muller TA, Hausinger RP (2008) Fe-II/alpha-ketoglutarate hydroxylases involved in nucleobase, nucleoside, nucleotide, and chromatin metabolism. *Dalton Trans*. pp 5132–5142.
- Shaffer PM, Hsu CA, Abbott MT (1975) Metabolism of pyrimidine deoxyribonucleosides in *Neurospora crassa*. *J Bacteriol* 121: 648–655.
- Palmarier RD, McCroskey RP, Abbott MT (1970) Enzymatic conversion of uracil 5-carboxylic acid to uracil and carbon dioxide. *J Biol Chem* 245: 6706–6710.
- Rusmintrapi V, Sowers LC (2000) An unexpectedly high excision capacity for mispaired 5-hydroxymethyluracil in human cell extracts. *Proc Natl Acad Sci USA* 97: 14183–14187.
- Boorstein RJ, Cummings A, Marenstein DR, Chan MK, Ma YL, et al. (2001) Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1. *J Biol Chem* 276: 41991–41997.
- Cannon SV, Cummings A, Teebor GW (1988) 5-Hydroxymethylcytosine DNA glycosylase activity in mammalian tissue. *Biochem Biophys Res Commun* 151: 1173–1179.
- Wossidlo M, Arand J, Sebastiano V, Lepikhov K, Boiani M, et al. (2010) Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. *EMBO J* 29: 1877–1888.
- Hajkova P, Jeffries SJ, Lee C, Miller N, Jackson SP, et al. (2010) Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* 329: 78–82.
- Wu SC, Zhang Y (2010) Active DNA demethylation: many roads lead to rome. *Nat Rev Mol Cell Biol* 11: 607–620.
- Feng J, Zhou Y, Campbell SL, Le T, Li E, et al. (2010) Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat Neurosci* 13: 423–430.
- Taghizadeh K, McFaline JL, Pang B, Sullivan M, Dong M, et al. (2008) Quantification of DNA damage products resulting from deamination, oxidation and reaction with products of lipid peroxidation by liquid chromatography isotope dilution tandem mass spectrometry. *Nature Protoc* 3: 1287–1298.
- Brückl T, Globisch D, Wagner M, Müller M, Carell T (2009) Parallel isotope-based quantification of modified tRNA nucleosides. *Angew Chem Int Ed* 48: 7932–7934.
- Yang Y, Nikolic D, Swanson SM, van Breemen RB (2002) Quantitative determination of N7-methyldeoxyguanosine and O6-methyldeoxyguanosine in DNA by LC-UV-MS-MS. *Anal Chem* 74: 5376–5382.

29. Hwang CH, Park JS, Won JH, Kim JN, Ryu EK (1992) The oxidative iodination of pyrimidine bases and their nucleosides using iodine/dimethylformamide/*m*-chloroperbenzoic acid. *Arch Pharm Res* 15: 69–72.
30. Gama-Sosa MA, Midgett RM, Slagel VA, Githens S, Kuo KC, et al. (1983) Tissue-specific differences in DNA methylation in various mammals. *Biochim Biophys Acta* 740: 212–219.
31. Vanyushin BF, Lopatina NG, Wise CK, Fullerton FR, Poirier LA (1998) Butylated hydroxytoluene modulates DNA methylation in rats. *Eur J Biochem* 256: 518–527.
32. Ming GL, Song HJ (2005) Adult neurogenesis in the mammalian central nervous system. *Annu Rev Neurosci* 28: 223–250.
33. Gage FH (2000) Mammalian neural stem cells. *Science* 287: 1433–1438.
34. Cooke MS, Olinski R, Loft S (2008) Measurement and meaning of oxidatively modified DNA lesions in urine. *Cancer Epidemiol Biomarkers Prev* 17: 3–14.

4.6 Postnatale, altersabhängige Änderungen der DNA-Modifikationslevel in Säugetieren

Ein umfassendes Verständnis der epigenetischen Bedeutung der Cytosinmodifikationen mC, hmC und fC ist ohne Kenntnis der dynamischen Änderungen der hmC- und mC-Level, welche innerhalb der Lebensspanne eines Säugetiers auftreten, kaum möglich. Schon in unserer ersten Publikation zum Thema Hydroxymethylcytosin (hmC) haben wir einen Anstieg des hmC- und Methylcytosin-Levels (mC) im Hippocampus von Mäusen innerhalb der ersten neunzig Tage nach der Geburt beobachtet. Ob es sich bei dieser Beobachtung um einen generellen Effekt handelt, der in verschiedenen Organen der Maus und vielleicht sogar bei allen Säugetieren auftritt, war zu diesem Zeitpunkt noch völlig unklar.

Ziel dieses Projektes war es daher, die postnatalen Dynamiken der modifizierten DNA-Nukleoside im Detail zu untersuchen. Neben murinen Geweben sollte diese Analyse auch menschliche Gewebe umfassen. Die dazu notwendige Quantifizierung der Basen sollte wieder mittels der in der Arbeitsgruppe Carell entwickelten LC-MS-basierten Isotopenverdünnungsmethode erfolgen. Mit dieser Untersuchung sollte der wissenschaftlichen Gemeinschaft erstmals ein verlässlicher Datensatz zu den altersabhängigen Dynamiken von mC, hmC und auch von fC im Menschen und in der Maus zur Verfügung gestellt werden.

In dieser Studie wurden die absoluten Level von hmC und mC in humanen Hirngeweben und von fC, hmC und mC in Cortex, Niere und Retina der Maus jeweils in Abhängigkeit vom Lebensalter bestimmt. Bei hmC wird mit zunehmendem Alter zunächst ein stetiger Anstieg der Modifikationsdichte beobachtet. Im voll entwickelten Gehirn steigt das hmC-Level nicht weiter an, sondern bleibt auf einem konstant hohen Niveau von ca. 1,2% beim Menschen. Dieser Wert ist in etwa doppelt so hoch wie das hmC-Level im zerebralen Cortex der Maus. Im Gegensatz dazu sinkt das fC-Level in den frühen Entwicklungsstadien und mit zunehmendem Alter stark ab. Insgesamt deuten diese Beobachtungen darauf hin, dass es sich bei hmC um eine stabile epigenetische Modifikation handelt, während fC in den frühen Gehirnentwicklungsphasen vermutlich eher als Intermediat der aktiven DNA-Demethylierung auftritt.

Die Ergebnisse dieses Projektes sind, z.T. auch während der Anfertigung dieser Dissertation, bereits publiziert worden. Da die Diskussion der Ergebnisse für die Publikationen erheblich gekürzt werden musste, habe ich mich entschieden, diesen Abschnitt dennoch in diese Arbeit mitaufzunehmen.

Postnatale, altersabhängige Änderungen der DNA-Modifikationslevel in Säugetieren

Schon in unserer ersten Publikation zum Thema Hydroxymethylcytosin (hmC) haben wir einen Anstieg des hmC- und Methylcytosin-Levels (mC) im Hippocampus von Mäusen innerhalb der ersten neunzig Tage nach der Geburt beobachtet. Wir konnten zeigen, dass dort das hmC-Level von 0,35% auf 0,6% und das mC-Level von 3,5% auf 4,5% steigt.^[113] Ob es sich bei dieser Beobachtung um einen generellen Effekt handelt, der in verschiedenen Organen der Maus und vielleicht sogar bei allen Säugetieren auftritt, war zu diesem Zeitpunkt noch völlig unklar. Mittlerweile sind die Änderungen der DNA-Modifikationsgrade während der Embryogenese und bei der Bildung der Urkeimzellen besser untersucht, z.T. sogar auf molekularer Ebene (siehe Einleitung).^[106, 150, 152, 155] Jedoch ist, von einigen Ausnahmen abgesehen, immer noch wenig über die Änderung der globalen hmC- und mC-Level während der weiteren Lebensspanne eines Organismus von der Geburt bis hin zum Tod bekannt.^[269, 282, 518-520]

Eine der wenigen Studien zu postnatalen, altersabhängigen Änderungen der DNA-Modifikationslevel in Säugetieren stammt von Szulwach *et al.*. Dort wurden die postnatalen, lokalen und globalen Änderungen des hmC-Modifikationsgrades im Hippocampus und im Cerebellum der Maus untersucht.^[269] Es wurde der hmC-Modifikationsgrad von sieben Tage (p7), sechs Wochen (pw6) und einem Jahr (pm12) alten Mäusen verglichen. Allerdings erfolgte hier die hmC-Quantifizierung über Immunhistochemie, so dass nur relative hmC-Level angegeben werden konnten. Ähnlich wie bei unseren ersten Messungen wurde in dieser Studie ein Anstieg des globalen hmC-Levels zwischen p7 und pm12 beobachtet. Im Hippocampus endet dieser Anstieg schon vor pw6, so dass Tiere im Alter von sechs Wochen bzw. einem Jahr ein vergleichbares hmC-Level aufweisen.

Auch Chen *et al.* haben die altersabhängigen Änderungen des hmC- und mC-Levels im Hippocampus der Maus untersucht.^[270] Das jeweilige hmC- und mC-Level wurde dabei durch einen Immunassay auf ELISA-Basis bestimmt. In dieser kurzen Studie konnte ein Anstieg des hmC-Levels um etwa 200% zwischen p2 und p14 beobachtet werden. Im Zeitraum zwischen zwei und 22 Monaten nach der Geburt (pm2-pm22) und auch zwischen pm4 und pm24 wurde ein Anstieg des hmC-Levels um nur noch etwa 30% bei gleichzeitig konstantem mC-Level ermittelt.

Von Lister *et al.* wurden im *Lobus frontalis* von Mensch und Maus die Änderungen des Methylierungsgrades auf molekularer Ebene über die gesamte Lebensspanne hinweg untersucht.^[282] Während der mittlere Methylierungsgrad der „klassischen“ mCG-Methylierungspositionen bei Mensch und Maus während der gesamten Lebensspanne praktisch konstant bleibt, wurden erhebliche altersabhängige Änderungen bei der nichtklassischen mCH-Methylierung beobachtet. An diesen

Methylierungspositionen findet beim Menschen von Geburt an bis ungefähr zum 16. Lebensjahr eine starke *de-novo*-Methylierung statt. Danach sinkt der Methylierungsgrad auf hohem Niveau bis zum Lebensende leicht ab. Bei der Maus beginnt diese *de-novo*-Methylierung bei p7 und ist bis pw4 sehr stark ausgeprägt. Das finale Methylierungslevel wird hier ungefähr sechs Wochen nach der Geburt (pw6 = p42) erreicht. Allerdings sind dann bei Mensch und Maus nur etwa 1,4% aller mCH-Positionen methyliert, während von den klassischen mCG-Positionen jederzeit ungefähr 80% methyliert sind.^[282]

Ein umfassendes Verständnis der epigenetischen Bedeutung der Cytosinmodifikationen hmC und mC ist ohne Kenntnis der dynamischen Änderungen der hmC- und mC-Level, welche innerhalb der Lebensspanne eines Säugetiers auftreten, kaum möglich. Von den oben beschriebenen Ausnahmen abgesehen, ist über diese postnatalen, altersabhängigen Änderungen der DNA-Modifikationslevel in Säugetieren bisher zu wenig bekannt.

In der Arbeitsgruppe Carell ist eine massenspektrometrie-basierte Isotopenverdünnungsmethode zur exakten Quantifizierung von modifizierten DNA-Basen in biologischen Proben entwickelt worden. Diese Methode ist ideal dafür geeignet, die Level der epigenetisch interessanten DNA-Modifikationen in biologischen Proben parallel zu bestimmen. Im Rahmen einer Kooperation mit der Arbeitsgruppe von Prof. Kretschmar habe ich deshalb zunächst die Änderung der hmC- und mC-Level in Abhängigkeit vom Lebensalter im menschlichen Gehirn untersucht. Die Ergebnisse dieses Projekts führten dann zu einer eingehenderen Untersuchung der altersbedingten Schwankungen der DNA-Modifikationslevel in verschiedenen Mäusegeweben. Bei diesen Untersuchungen, welche ich zusammen mit Jessica Steinbacher durchgeführt habe, konnte auch die kurz zuvor von Carell und Mitarbeitern entdeckte DNA-Base Formylcytosin (fC) in verschiedenen Geweben quantifiziert werden.^[25] Später habe ich dann die altersabhängigen Änderungen dieser Nukleosidlevel im Detail zusammen mit der Arbeitsgruppe Michalakakis in der Retina von Mäusen studiert.

4.6.1 Altersabhängige Variation der hmC- und mC-Level im menschlichen Hirn

Von Prof. Kretschmar wurden mir Proben aus dem Gehirn von sieben verschiedenen Patienten zur Verfügung gestellt. Zum Zeitpunkt der Probenentnahme lag das Alter der Patienten zwischen der 15. Schwangerschaftswoche (15. SSW) und 85 Jahren (85 a), wobei es in der Probenreihe aufgrund der natürlichen Sterberate eine größere zeitliche Lücke zwischen 3,2 und 22 Jahren gibt. Von allen Patienten habe ich jeweils eine Probe aus der grauen und weißen Substanz des Cerebrums (d.h. des Großhirns) erhalten. Die graue Substanz (Rinde) enthält dabei im Vergleich zur weißen Substanz (Mark) eine höhere Anzahl von Zellkörpern und nur relativ wenige myelinisierte Axone. Die Proben

wurden entweder am *Lobus frontalis* (Frontallappen) oder am *Lobus occipitalis* (Okzipitallappen bzw. Hinterhauptslappen) entnommen (Abbildung 31).

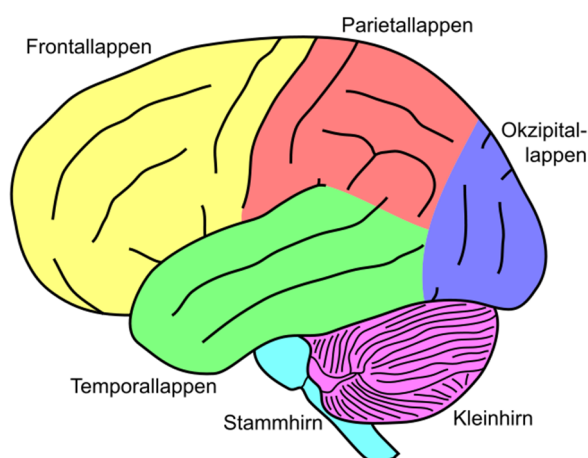


Abbildung 31. Schematische Darstellung des menschlichen Gehirns.^[521]

Der *Lobus frontalis* enthält den Motorcortex und gilt auch als Sitz der Persönlichkeit und des Sozialverhaltens, während der Großteil des *Lobus occipitalis* vom visuellen Cortex (der Sehrinde) eingenommen wird. Von zwei Patienten (22 a und 85 a) habe ich zusätzlich Proben von grauer und weißer Substanz des Cerebellums (d.h. des Kleinhirns) erhalten.

Nach Extraktion der DNA, welche nach der von Carell und Mitarbeitern entwickelten Methode erfolgte,^[18, 113] konnten die Modifikationen hmC und mC in allen Proben erfolgreich massenspektrometrisch (HPLC-ESI-MS) quantifiziert werden. Die oben erwähnte, hauptsächlich von Daniel Globisch in der Arbeitsgruppe Carell entwickelte Isotopenverdünnungsmethode erlaubte dabei die Bestimmung der absoluten Nukleosidlevel in den Proben.^[18] Diese Methode basiert auf der Zugabe (dem sog. Spiking) von stabilen, isotonenmarkierten Nukleosid-Standards zur Probe. Die Massenspektrometrie ist zwar eine extrem empfindliche Analysemethode, die auch den Nachweis kleinster Stoffmengen erlaubt, die Messergebnisse *per se* sind jedoch nicht quantitativ bzw. absolut. Erst die Verwendung von stabil isotonenmarkierten Referenznukleosiden, welche zuvor von Carell und Mitarbeitern synthetisiert werden mussten, ermöglicht in der hier verwendeten Methode die Bestimmung der absoluten Nukleosidlevel. Dabei besitzen die isotonenmarkierten Standards dasselbe chromatographische Verhalten und dieselben Ionisationseigenschaften wie die natürlichen Nukleoside, sie haben jedoch ein anderes Molekulargewicht. Dies erlaubt eine separate Integration der Massensignale der natürlichen Nukleoside, deren Menge quantifiziert werden soll, und der isotonenmarkierten Standards, deren Menge in der Probe bekannt ist. Der Vergleich der

entsprechenden Integrale unter Verwendung einer Eichgeraden erlaubt dann eine äußerst präzise, absolute Quantifizierung der natürlichen Nukleoside in der Probe.

4.6.1.1 Dynamik der Hydroxymethylcytosinlevel im menschlichen Hirn

Betrachtet man die Ergebnisse der HPLC-ESI-MS-Messungen (Abbildung 32), welche hier in Prozent pro dG angegeben werden, so fällt als erstes der starke altersabhängige Anstieg des hmC-Level während der Entwicklung vom Fötus zum erwachsenen Menschen auf. Dieser Anstieg ist sowohl im Cerebrum als auch im Cerebellum zu beobachten und betrifft sowohl die graue, als auch die weiße Substanz.

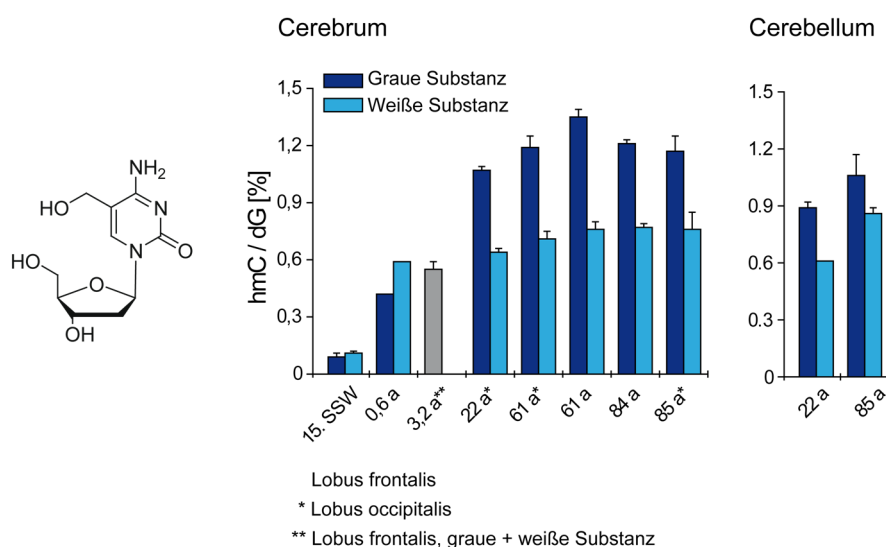


Abbildung 32. Dynamik der Hydroxymethylcytosinlevel im menschlichen Gehirn in Abhängigkeit vom Lebensalter. Die Fehlerbalken ergeben sich aus der Standardabweichung bei (je nach Probenmenge) zwei bis vier technischen Replikaten. Beim 0,6 Jahre alten Patienten konnte wegen der geringen Probenmenge nur eine Messung durchgeführt werden. SSW = Schwangerschaftswoche.

Im Cerebrum liegt das hmC-Level in der 15. Schwangerschaftswoche (SSW) beim Fötus nur bei 0,09 bzw. 0,11% (graue bzw. weiße Substanz, d.h. Rinde bzw. Mark). Es steigt dann mit zunehmendem Alter von 0,42 / 0,59% (0,6 a) über 0,55% (3,2 a; graue + weiße Substanz) auf 1,07 / 0,64% bei einem 22 Jahre alten Patienten stark an. In der grauen Substanz steigt das hmC-Level danach nur noch wenig an und stagniert ab einem Alter von etwa 60 Jahren auf hohem Niveau. Die deutliche Verlangsamung der hmC-Zunahme setzt in der weißen Substanz schon früher ein. Der genaue Zeitpunkt ist schwer zu bestimmen, liegt aber zwischen dem ersten und dem zweiundzwanzigsten Lebensjahr. Patienten im Alter von 61 (2x), 84 und 85 Jahren besitzen daher ein ähnliches hmC-Level von durchschnittlich $1,23 \pm 0,07\%$ in der grauen bzw. $0,75 \pm 0,02\%$ in der weißen Substanz. Ein

ähnlicher zeitlicher Verlauf des hmC-Modifikationsgrades wurde von Szulwach *et al.* im Hippocampus der Maus beobachtet. Auch hier kommt es in der ersten Lebensphase zu einem starken Anstieg des hmC-Levels. Ab dem frühen Erwachsenenalter steigt das hmC-Level nicht weiter an, sondern bleibt während der restlichen Lebensspanne der Maus auf konstant hohem Niveau.^[269]

Interessanterweise kommt es in einem Alter von ungefähr drei Jahren zu einer Inversion der absoluten hmC-Level von grauer und weißer Substanz im Cerebrum: Während das hmC-Level in der weißen Substanz noch beim Fötus und im Alter von 0,6 Jahren höher als in der grauen Substanz ist, kehrt sich dieses Verhältnis deutlich vor dem zweiundzwanzigsten Lebensjahr um. Von diesem Zeitpunkt an beträgt das Verhältnis von hmC in grauer und weißer Substanz im Durchschnitt 1,6 zu 1.

Von dem 22- und 85-jährigen Patienten konnten auch Proben des Cerebellums erhalten und analysiert werden. Auch hier wurde ein altersabhängiger Anstieg des hmC-Levels beobachtet. Dabei steigt das hmC-Level von 0,89 auf 1,06% in der grauen und von 0,61 auf 0,86% in der weißen Substanz. Im Vergleich zu den entsprechenden Messwerten im Cerebrum sind die im Cerebellum detektierten Werte nur unbedeutend kleiner.

4.6.1.2 Dynamik der Methylcytosinlevel im menschlichen Hirn

Auch beim mC ist ein altersabhängiger Anstieg des DNA-Modifikationsgrades zu beobachten (Abbildung 33). Dieser tritt jedoch nur im Cerebrum und dort besonders in der grauen Substanz auf. Er ist außerdem deutlich geringer ausgeprägt als der altersabhängige Anstieg des hmC-Levels in denselben Proben. In der grauen Substanz steigt das mC-Level von 3,7% (Fötus in der 15. Schwangerschaftswoche) über 4,4% (0,6 a) auf 5,9% (22 a). Offenbar endet der Anstieg des mC-Levels etwa mit dem zweiundzwanzigsten Lebensjahr und damit früher als die beobachtete hmC-Zunahme. Ab diesem Alter nimmt der Methylierungsgrad auf hohem Niveau wieder leicht ab. Die Patienten ab einem Alter von 61 Jahren weisen dann einen ähnlichen DNA-Methylierungsgrad in der grauen Substanz von durchschnittlich $5,2 \pm 0,1\%$ auf.

Unsere Beobachtungen stimmen gut mit dem von Lister *et al.* beobachteten zeitlichen Verlauf der *de-novo*-Methylierung von nichtklassischen (d.h. nicht-mCG) mCH-Methylierungspositionen in der DNA überein.^[282] Hier wird der maximale Methylierungsgrad im *Lobus frontalis* des Menschen im Alter von etwa 16 Jahren erreicht. Möglicherweise ist auch der von uns in der grauen Substanz beobachtete Anstieg des globalen mC-Levels hauptsächlich auf solche „nichtklassischen“ *de-novo*-Methylierungen zurückzuführen.

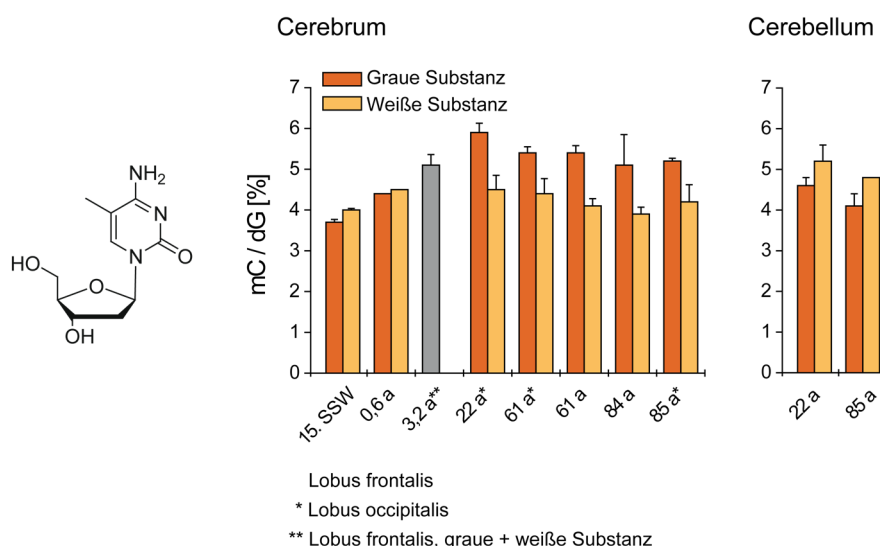


Abbildung 33. Dynamik der Methylcytosinlevel im menschlichen Gehirn in Abhängigkeit vom Lebensalter. Die Fehlerbalken ergeben sich aus der Standardabweichung bei (je nach Probenmenge) zwei bis vier technischen Replikaten. Beim 0,6 Jahre alten Patienten konnte wegen der geringen Probenmenge nur eine Messung durchgeführt werden. SSW = Schwangerschaftswoche.

Das postnatale mC-Level in der weißen Substanz variiert dagegen praktisch gar nicht. Es liegt hier während der gesamten Lebensspanne bei ungefähr $4,2 \pm 0,2\%$. Ab dem 22. Lebensjahr ist hier eher eine leichte Abnahme des Methylierungsgrades zu beobachten.

Auch bei mC sind die in der weißen Substanz des Cerebrums gemessenen Level beim Fötus und im Alter von 0,6 Jahren etwas höher als in der grauen Substanz. Analog zu hmC kehrt sich dieses Verhältnis vor dem zweiundzwanzigsten Lebensjahr um. Das Verhältnis beträgt dann im Durchschnitt 1,3 zu 1, so dass der Unterschied im Methylierungsgrad von grauer und weißer Substanz nicht so groß wie bei hmC ist. Im Cerebellum findet man dagegen überraschenderweise auch bei dem 22- und 85-jährigen Patienten mehr mC in der weißen als in der grauen Substanz.

4.6.2 Altersabhängige Änderungen der Nukleosidlevel in Mäusegeweben

Unsere neuen Erkenntnisse über die altersabhängige Zunahme von hmC im menschlichen Gehirn passten sehr gut zu unseren bereits publizierten hmC-Quantifizierungen in Mäusegeweben. Dort hatten wir im Hippocampus einen Anstiegs des hmC-Levels innerhalb der ersten neunzig Tage nach der Geburt beobachtet.^[113] Um die Frage zu beantworten, ob dieser Effekt auch in anderen Organen auftritt, haben wir uns für eine weitere Untersuchung des Alterseffektes wieder der Maus als klassischem Säugetier-Modellorganismus zugewandt.

4.6.2.1 Altersabhängige Änderungen der Nukleosidlevel in Cerebrum und Niere

Um einen direkten Vergleich mit den humanen Daten zu ermöglichen, haben Jessica Steinbacher und ich den zeitlichen Verlauf des hmC- und mC-Levels im Cerebrum der Maus untersucht. Dazu wurden Proben des zerebralen Cortex (d.h. der Großhirnrinde) analysiert. Auf Basis der bereits von uns publizierten Quantifizierungsdaten von 90 Tage alten Mäusen wurde daneben die Niere als nicht-neuronales Gewebe für das weitere Studium des Alterseffektes ausgewählt.^[18] Die Niere ist dabei ein typischer Vertreter von Organen mit mittlerem hmC-Modifikationsgrad von ca. 0,2%. Dieser Modifikationsgrad ist vergleichbar mit dem in Herz, Lunge und Muskeln detektierten hmC-Level.^[31]

Für diese Untersuchungen wurde eine neue, UHPLC-ESI-MS/MS-basierte Analysemethode verwendet. Diese zum Großteil von Toni Pfaffeneder entwickelte Methode erlaubt im Vergleich zu der bei den humanen Proben verwendeten Methode eine sensitivere Detektion der Nukleosidlevel.^[31]

Die Untersuchungen wurden an weiblichen Wildtyp-Mäusen (C57-BL6/N) im Alter von einem Tag, 14 Tagen und 90 Tagen bzw. drei Monaten durchgeführt (p1, p14 und p90 = pm3); p = postnatal, d.h. Tage nach der Geburt). Bei drei Monate alten Mäusen handelt es sich bereits um adulte Tiere im frühen Erwachsenenalter (Mäuse sind sechs Wochen nach der Geburt geschlechtsreif und werden in Gefangenschaft durchschnittlich 2-3 Jahre alt).

Bei p14 und p90 wurden jeweils drei biologische Replikate analysiert, bei p1 konnten aufgrund eines technischen Problems nur von jeweils zwei biologischen Replikaten Daten erhalten werden. Je Probe wurden drei Messungen durchgeführt (technische Replikate).

Dynamik der Hydroxymethyl- und Methylcytosinlevel

Betrachtet man die Ergebnisse der hmC-Quantifizierungen (Abbildung 34), welche hier in Prozent pro dG angegeben werden, so findet man auch im Cerebrum und in der Niere den bereits im murinen Hippocampus und beim Menschen beobachteten altersabhängigen Anstieg des hmC-Levels. Wie erwartet, sind dabei die im Cerebrum detektierten hmC-Level deutlich größer als in der Niere: Im Cerebrum steigt das hmC-Level von 0,16% (p1) über 0,24% (p14) auf 0,57% (p90). In der Niere beobachtet man im selben Zeitraum einen Anstieg von 0,05% (p1) über 0,08% (p14) auf 0,18% (p90). Der Anstieg des hmC-Levels von p1 nach p12 beträgt damit in beiden Geweben ca. 250%. Dieser Wert passt gut zu dem von Chen *et al.* im Hippocampus der Maus zwischen p2 und p14 gemessenen Anstieg des hmC-Levels um ca. 200%.^[270]

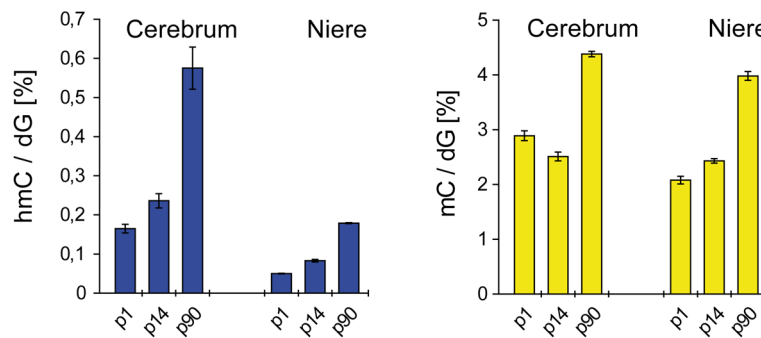


Abbildung 34. Altersabhängige Änderung des Hydroxymethyl- und Methylcytosinlevels in Cerebrum (zerebraler Cortex) und Niere der Maus zwischen p1 und p90 (p = postnatal, d.h. Tage nach der Geburt). Gemessen wurden Proben von weiblichen Wildtyp-Mäusen (C57-BL6/N). Die Fehlerbalken ergeben sich aus der Standardabweichung bei zwei (p1) bzw. drei (p14 und p90) biologischen Replikaten.

Vergleicht man das hmC-Level im adulten Cerebrum von Mensch (22 a) und Maus (p90), so zeigt sich, dass der Hydroxymethylierungsgrad beim Menschen (1,07% in der grauen bzw. 0,64% in der weißen Substanz) in der grauen Substanz größer ist als bei der Maus (0,57%). Eine Erklärung für diesen Unterschied liefert möglicherweise die unterschiedliche Probenentnahme: Die Proben enthielten bei der Maus einen Großteil des Cerebrums. Die graue Substanz, in der beim Menschen besonders viel hmC gefunden wurde, nimmt jedoch nur einen kleinen Teil des Cerebrums ein. Der Anteil an weißer Substanz, die bei Mensch und Maus ein ähnliches hmC-Level aufweist, ist dagegen deutlich größer.

Die in Cerebrum und Niere der Maus gemessenen mC-Level sind besonders zu den Zeitpunkten p14 und p90 ähnlich groß (Abbildung 34). Darüber hinaus findet man in beiden Organen einen starken altersabhängigen Anstieg des DNA-Methylierungsgrades um ungefähr 70% zwischen p14 und p90. Zwischen p1 und p14 ändert sich das mC-Level in Cerebrum und Niere dagegen nur wenig. Hier beobachtet man im Cerebrum einen leichten Rückgang des mC-Levels von 2,9% auf 2,5%, welcher sich anhand der von uns gemessenen hmC-Level nicht durch eine Oxidation von mC zu hmC erklären lässt. In der Niere steigt das mC-Level in diesem Zeitraum dagegen von 2,1% auf 2,4% leicht an.

Auch von Lister *et al.* wurde der zeitliche Verlauf des Methylierungsgrades im Cerebrum der Maus (und dort im frontalen Cortex) untersucht, allerdings auf molekularer Ebene.^[282] Ähnlich wie wir haben Lister *et al.* einen verzögerten, dann aber starken Anstieg des Methylierungsgrades im Cerebrum der Maus beobachtet, der ungefähr bei p7 beginnt und bei ca. pw6 (= p42) endet. Während der restlichen Lebensspanne ab pw6 wurde von Lister *et al.* eine Stagnation des mC-Levels auf diesem hohen Niveau beobachtet. Daher entsprechen die von uns bei p90 gemessenen globalen mC-Level wahrscheinlich zumindest im Cerebrum diesem von Lister beobachteten *steady-state*-Level des Methylierungsgrades in der späteren Lebensphase. Die von Lister *et al.* gefundenen *de-novo*-

Methylierungen finden bei Maus und Mensch fast ausschließlich an nichtklassischen (d.h. nicht-mCG) mCH-Methylierungspositionen der DNA statt.^[282] Möglicherweise ist daher auch der von uns in Cerebrum und Niere der Maus beobachtete Anstieg des globalen mC-Levels hauptsächlich auf solche „nichtklassischen“ *de-novo*-Methylierungen zurückzuführen.

Ähnlich wie bei hmC gleicht das mC-Level im adulten murinen Cerebrum (4,4%, p90) dem mC-Level in der weißen Substanz des humanen Cerebrums (4,5%, 22 a). In der grauen Substanz des humanen Cerebrums findet man dagegen einen höheren Methylierungsgrad (5,9%, 22 a) (Abbildung 33). Der im Cerebrum der Maus beobachtete Anstieg des mC-Levels um ca. 50% zwischen p1 und p90 fällt etwas stärker aus als im humanen Cerebrum. In der grauen Substanz steigt dort der Methylierungsgrad im analogen Zeitraum zwischen 0,6 und 22 Jahren nur um etwa 30% an, während sich das mC-Level in der weißen Substanz in diesem Zeitraum nicht ändert.

Dynamik des Formylcytosinlevels

Für die in hier beschriebene Quantifizierung der Nukleosidlevel in Mäusegeweben nach der neuen UHPLC-ESI-MS/MS-Methode wurde von uns ein äußerst sensitives Triple-Quadrupol-Massenspektrometer verwendet. Dies erlaubte prinzipiell auch die Quantifizierung der DNA-Modifikationen Formylcytosin (fC), welche erst kurz zuvor von Carell und Mitarbeitern in embryonalen Stammzellen entdeckt worden war, und Carboxycytosin (caC).^[25] Tatsächlich konnten wir die Base fC in allen hier untersuchten Proben genau quantifizieren. Jedoch konnte trotz intensiver Suche in keinem der hier analysierten Gewebe die Base caC gefunden werden.

Auch beim fC beobachtet man in Cerebrum und Niere der Maus eine starke altersabhängige Änderung des Modifikationsgrades zwischen p1 und p90 (Abbildung 35). Im Gegensatz zu hmC und mC kommt es hier jedoch zu einer Abnahme DNA-Modifikationsgrades. Dieser Rückgang ist umso bemerkenswerter, da sowohl hmC als auch fC durch TET-vermittelte Oxidation von mC gebildet werden.^[23, 26] Wegen der gegenläufigen Trends bei mC und hmC einerseits und fC andererseits kommt es zu einer starken Änderung der relativen Häufigkeit dieser Basen zwischen p1 und p90. Im Cerebrum beträgt das Verhältnis mC : hmC : fC 4.400 : 250 : 1 (p1) bzw. 22.000 : 2.900 : 1 (p90). In der Niere findet man ein Verhältnis von 3.200 : 77 : 1 (p1) bzw. 44.000 : 2.000 : 1 (p90).

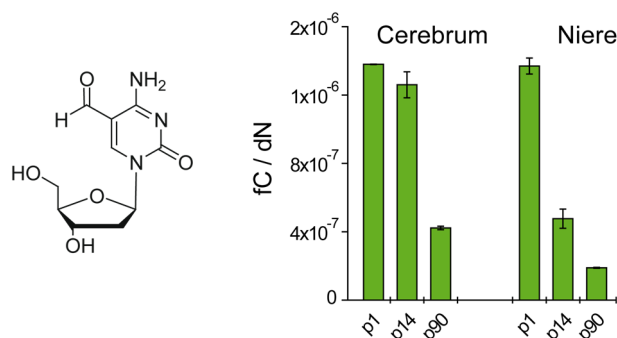


Abbildung 35. Altersabhängige Änderung des Formylcytosinlevels in Cerebrum (zerebraler Cortex) und Niere der Maus zwischen p1 und p90 (p = postnatal, d.h. Tage nach der Geburt). Gemessen wurden Proben von weiblichen Wildtyp-Mäusen (C57-BL6/N). Die Fehlerbalken ergeben sich aus der Standardabweichung bei zwei (p1) bzw. drei (p14 und p90) biologischen Replikaten.

Der Rückgang des fC-Levels setzt im Cerebrum später als in der Niere ein. So ist der DNA-Formylierungsgrad im Cerebrum einen und 14 Tage nach der Geburt mit $1,4 \times 10^{-6}$ bzw. $1,3 \times 10^{-6}$ fC-Basen pro dN fast gleich hoch. Erst zwischen p14 und p90 findet ein starker Rückgang des fC-Levels um etwa 65% statt, so dass man im adulten Tier (p90) nur noch $4,2 \times 10^{-7}$ fC-Basen pro dN findet.

In der Niere findet man einen Tag nach der Geburt mit $1,4 \times 10^{-6}$ fC-Basen pro dN noch genauso viel fC wie im Cerebrum. Während das im Cerebrum gemessene fC-Level bis p14 fast konstant bleibt, beobachtet man in diesem Zeitraum in der Niere eine Abnahme des fC-Levels um 65% auf $4,8 \times 10^{-7}$ fC-Basen pro dN. Zwischen p14 und p90 kommt es zu einem weiteren Rückgang des fC-Levels um 60%, so dass man im adulten Tier nur noch $1,9 \times 10^{-7}$ fC-Basen pro dN findet. Insgesamt ist daher der Formylierungsgrad im Cerebrum bei der adulten Maus (p90) etwa doppelt so hoch wie in der Niere.

Interessanterweise sind die fC-Level in Cerebrum und Niere kurz nach der Geburt besonders hoch. Zu diesem Zeitpunkt sind die Organe, besonders das Gehirn, noch in der Entwicklung begriffen. Wie mittlerweile von vielen Gruppen gezeigt werden konnte, besitzt fC besonders als Intermediat der aktiven DNA-Demethylierung eine große Bedeutung (siehe Einleitung).^[34, 103] Geht man davon aus, dass kurz nach der Geburt eine besonders starke epigenetische Remodellierung, u.a. durch aktive DNA-Demethylierung, in den noch in der Entwicklung befindlichen Organen stattfindet, die mit fortlaufender Differenzierung der Organe abnimmt, so würde dies eine gute Erklärung für die von uns beobachtete altersabhängige Abnahme der fC-Level liefern. In diesem Fall würde es sich beim Großteil des von uns detektierten fCs um transiente Formylierungen handeln; das fC-Level wäre dann ein guter Indikator für die momentane Demethylierungsaktivität im entsprechenden Organ.

Folgt man dieser Argumentation, so liefern die hier gemessenen fC-Level auch eine mögliche Erklärung für den im Cerebrum beobachteten leichten Rückgang des mC-Levels zwischen p1 und p14 (Abbildung 34). In diesem Fall würde das 14 Tage nach der Geburt gemessene, im Vergleich zur Niere deutlich höhere fC-Level eine zu diesem Zeitpunkt immer noch starke Demethylierungsaktivität im Cerebrum widerspiegeln. Überwiegt diese Demethylierungsaktivität die *de-novo*-Methylierungen, so folgt daraus der beobachtete leichte Rückgang des mC-Levels.

4.6.2.2 Detaillierte Untersuchung der altersabhängigen Änderungen der Nukleosidlevel in der Retina

Die im menschlichen Gehirn und in der Maus beobachteten altersabhängigen Änderungen der Modifikationsgrade von mC, hmC und fC sollten in einem nachfolgenden Projekt *en détail* studiert werden. Um auf eine möglichst breite Datenbasis zurückgreifen zu können und wegen der leichten Probenzugänglichkeit bot sich auch in diesem Fall die Maus als Modellorganismus an. Die Retina wurde von mir als Modellorgan gewählt, weil dies eine Kooperation mit den Retinaspezialisten der Arbeitsgruppe Michalakakis mit einem entsprechenden Wissenstransfer erlaubte. Außerdem handelt es sich auch bei der Retina um ein neuronales Gewebe mit einem relativ hohen hmC-Level.^[113]

Um besonders die erwartete Abnahme des fC-Levels im Detail untersuchen zu können, wurden von mir Retinaprobe analysiert, die 2, 10, 12, 13 und 21 Tage nach der Geburt der Mäuse entnommen worden waren (p2-p21). Die Häufung der Probenentnahme um p12 herum ergab sich aus dem Interessenschwerpunkt des Kooperationspartners. Die Arbeitsgruppe von Dr. Michalakakis war in dieser Kooperation besonders daran interessiert herauszufinden, welche Änderungen in der Retina stattfinden, wenn eine Maus zum ersten Mal in ihrem Leben die Augen öffnet. (Mäuse werden mit geschlossenen Augen geboren und öffnen diese zum ersten Mal ca. 12–13 Tage nach der Geburt.)

Auch in der murinen Retina wurden die Cytosinmodifikationen mit unserer UHPLC-ESI-MS/MS-Methode unter Verwendung eines Triple-Quadrupol-Massenspektrometers quantifiziert. Es wurden jeweils drei biologische Replikate analysiert. Eine Ausnahme bildete p12, wo aufgrund eines technischen Fehlers nur von zwei biologischen Replikaten Daten erhalten werden konnten. Je Probe wurden drei Messungen durchgeführt (technische Replikate).

Dynamik des mC-, hmC- und fC-Levels in der Retina

Wie ausgehend von unseren bisherigen Daten zu erwarten war, beobachtet man auch in der Retina nach der Geburt einen starken Anstieg des hmC-Levels (Abbildung 36). Das hmC-Level steigt hier

zwischen p2 und p21 kontinuierlich von 0,07% auf 0,21% an. Allerdings nimmt der jeweilige Anstieg des hmC-Level mit wachsendem Alter immer weiter ab. Die retinalen hmC-Level zu den Zeitpunkten p2 und p13, welche mit den in Cerebrum und Niere bei p1 bzw. p14 gemessenen Werten verglichen werden können, liegen zwischen den bei diesen Geweben beobachteten Leveln.

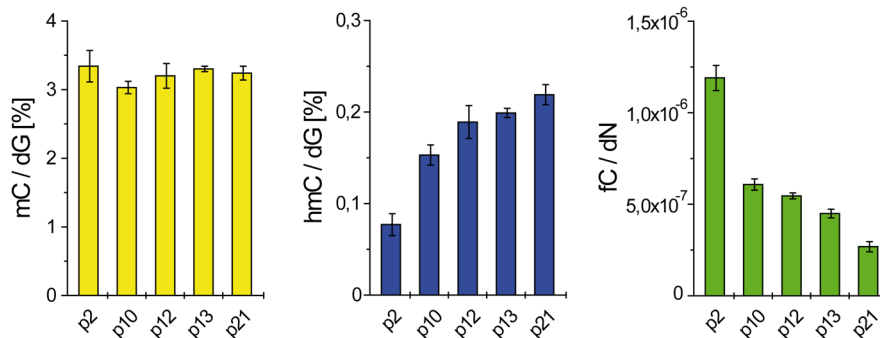


Abbildung 36. Altersabhängige Änderung des mC-, hmC- und fC-Levels in der murinen Retina zwischen p2 und p21 (p = postnatal, d.h. Tage nach der Geburt). Analysiert wurden Retinae von weiblichen Wildtyp-Mäusen (C57-BL6/N). Die Fehlerbalken ergeben sich aus der Standardabweichung bei zwei (p12) bzw. drei (p2, 10, 13 und 21) biologischen Replikaten.

Im Gegensatz zu dem bei hmC beobachteten Anstieg ändert sich das mC-Level in der Retina zwischen p2 und p21 nicht, sondern bleibt konstant bei $3,0 \pm 0,1\%$. (Die im Cerebrum und in der Niere der Maus beobachtete starke Zunahme des DNA-Methylierungsgrades zwischen p14 und p90 wird von unseren Retinaprobe nicht abgedeckt, hier endet die Zeitreihe bei p21.) Würde das von uns in der Retina detektierte hmC hauptsächlich als Zwischenstufe einer aktiven DNA-Demethylierung dienen, so dass ein hohes hmC-Level Zeichen einer starken Demethylierungsaktivität wäre, dann würde man zumindest eine leichte Abnahme des DNA-Methylierungsgrades zwischen p2 und p21 erwarten. Daher passt das hier beobachtete konstante mC-Level und der in Cerebrum und Niere zwischen p14 und p90 beobachtete Anstieg des Methylierungsgrades bei gleichzeitiger Zunahme von hmC (Abbildung 34) gut zu der inzwischen allgemein akzeptierten Theorie, dass es sich bei hmC um eine relativ stabile Modifikation mit eigener epigenetischer Bedeutung handelt.^[40, 278, 290] Allerdings hätte man in der Retina spätestens bei p21 den von Lister *et al.* im murinen Cerebrum zwischen p7 und p28 beobachteten starken Anstieg des Methylierungsgrades erwartet, da es sich auch bei der Retina um ein neuronales Gewebe handelt.^[282]

Auch in der Retina beobachtet man beim fC eine Abnahme des Modifikationslevels in Abhängigkeit vom Alter (Abbildung 36). So findet man in der Retina von 2 Tage alten Mäusen (p2) $1,2 \times 10^{-6}$ fC-Basen pro dN. Zehn Tage nach der Geburt (p10) findet man nur noch $5,8 \times 10^{-7}$ fC-Basen pro dN.

Dieser Trend setzt bei allen untersuchten Altersstufen fort (p12, p13, p21), bis das fC-Level 21 Tage nach der Geburt nur noch $2,8 \times 10^{-7}$ Basen pro dN beträgt.

Durch die Abnahme von fC bei gleichzeitiger Zunahme von hmC kommt es zu einer Änderung des Verhältnisses der beiden Modifikationen zueinander. Berücksichtigt man auch die Base mC, so findet man zwei Tage nach der Geburt in der Retina ein Verhältnis mC : hmC : fC von 5.800 : 125 : 1. Einundzwanzig Tage nach der Geburt beträgt dieses Verhältnis dann 23.900 : 1.600 : 1.

Natürlich haben wir auch in den Retinaprobe nach der DNA-Modifikation Carboxycytosin (caC) gesucht, es konnte jedoch trotz intensiver Suche wie in den anderen Mäusegeweben kein caC detektiert werden.

4.6.2.3 Analyse der in der Retina gemessenen Nukleosidlevel anhand der Expressionsdaten wichtiger DNA-modifizierender Enzyme

Die Expressionslevel wichtiger DNA-modifizierender Enzyme waren in der Retina bereits von Arshan Perera aus der Arbeitsgruppe Michalakakis über quantitative PCR bestimmt worden. U.a. standen uns damit Expressionsdaten der drei TET-Enzyme zur Verfügung. Daneben waren auch das Expressionslevel der Glykosylase TDG, welche fC und caC aus der DNA herausschneiden kann,^[27, 34, 350] und die Expressionslevel der DNA-Methyltransferasen DNMT1 und DNMT3a bestimmt worden (**Abbildung 37**). Die für die Bestimmung der Expressionslevel gewählten Zeitpunkte (p2, 11, 12, 13 und 20) stimmen zwar nicht immer exakt mit den von uns gewählten Zeitpunkten der Probenentnahme (p1, 10, 12, 13 und 21) überein, die zeitliche Nähe ist aber hinreichend groß, um eine Interpretation der von uns gemessenen Nukleosidlevel zu ermöglichen.

Betrachtet man die die Expressionslevel der drei TET-Enzyme, so fällt auf, dass TET3 während des gesamten Zeitraums zwischen p2 und p20 mit Abstand am stärksten exprimiert wird. Dagegen sind die Expressionslevel von TET1 und TET2 auf niedrigem Niveau jeweils vergleichbar groß. Besonders auffallend ist ein starker Abfall der Expressionslevel bei allen drei TET-Enzymen zwischen p12 und p13. Obwohl die TET- Enzyme für die Generierung von hmC und fC verantwortlich sind, spiegelt sich diese starke Verminderung der TET-Expression interessanterweise nicht in den globalen hmC- bzw. fC-Leveln der Retina.^[23, 109] Stattdessen beobachtet man, wie oben beschrieben, in diesem Zeitraum auf Nukleosidebene eine kontinuierliche Zu- bzw. Abnahme der Modifikationsdichte bei hmC bzw. fC (Abbildung 36). Der Abfall der TET-Expressionslevel zwischen p12 und p13 fällt jedoch genau mit dem Zeitpunkt der ersten Augenöffnung bei der Maus zusammen. Die hier praktizierte globale Quantifizierung der Nukleosidlevel erlaubt keine Rückschlüsse auf subtile lokale Änderungen des DNA-Modifikationsgrades. Vermutlich sind es jedoch hauptsächlich solche subtilen Änderungen, die

die in der Retina auf molekularer und makroskopischer Ebene zum Zeitpunkt der ersten Augenöffnung beobachteten Veränderungen steuern. Dies würde erklären, warum sich die mit der ersten Augenöffnung vermutlich einhergehende epigenetische Reprogrammierung nicht in den von uns beobachteten Nukleosidleveln widerspiegelt.

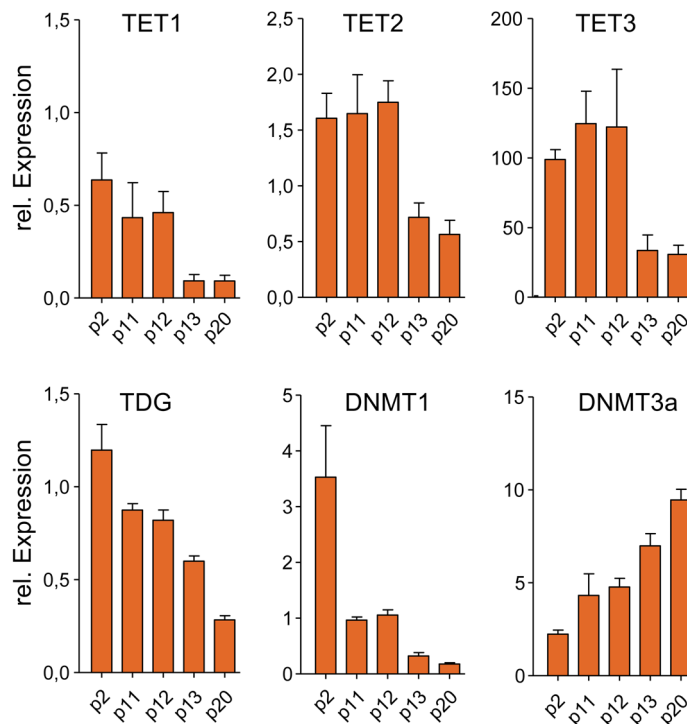


Abbildung 37. Expressionslevel wichtiger DNA-modifizierender Enzyme in der murinen Retina zwischen p2 und p20, ermittelt durch quantitative PCR.

Der beobachtete Anstieg des hmC-Levels bei gleichzeitiger Abnahme der TET-Expression lässt sich mit der allgemein akzeptierten Theorie erklären, dass es sich bei hmC um eine relativ stabile Modifikation mit eigener epigenetischer Bedeutung handelt:^[40, 278, 290] Die hohe TET-Expression kurz nach der Geburt sorgt möglicherweise für den beobachteten Anstieg des hmC-Levels. Da hmC offenbar eine hohe „Lebensdauer“ hat und nicht sofort weiteroxidiert oder aus der DNA herausgeschnitten wird, kommt es bei einer verminderten TET-Expression nicht zu einem Absinken des hmC-Levels. Stattdessen kommt es zu der beobachteten, allerdings stark verlangsamten hmC-Akkumulation. Nimmt die TET-Expression zu sehr ab, würde es nach dieser Theorie zu einer Stagnation des hmC-Levels auf hohem Niveau kommen, wie wir sie im humanen Cerebrum beobachtet haben und wie sie von Szulwach *et al.* im Hippocampus der Maus gefunden wurde.^[269]

Bei TDG beobachtet man eine kontinuierliche Abnahme des Expressionslevels von p2 bis p20 um ca. 80%. Diese Abnahme korreliert sehr gut mit der beobachteten Abnahme des fC-Levels im selben

Zeitraum um 75%. Man würde intuitiv erwarten, dass eine starke TDG-Expression direkt zu einem besonders niedrigen fC-Level auf DNA-Ebene führen sollte. Stattdessen ist es möglicherweise so, dass es sich bei den von uns in der DNA gefundenen fC-Basen hauptsächlich um Intermediate der TDG-vermittelten aktiven DNA-Demethylierung handelt.^[34] Das hohe retinale fC-Level in den ersten Tagen nach der Geburt würde dann auf eine stark erhöhte Demethylierungsaktivität in diesem Zeitraum hinweisen. Da sich die Retina zu diesem Zeitpunkt noch in der Entwicklung befindet, ist das Vorhandensein einer starken epigenetischen Remodellierung, möglicherweise auch durch aktive DNA-Demethylierung, nicht unwahrscheinlich. Wenn man davon ausgeht, dass das Expressionslevel von TDG reguliert und dem aktuellen Bedarf im Gewebe angepasst ist, würde dies das hohe TDG-Expressionsniveau zum selben Zeitpunkt erklären. Das würde allerdings auch bedeuten, dass die fC Level vermutlich deutlich höher sind, als von uns detektiert (da ein großer Teil des fCs durch TDG kontinuierlich und direkt nach der Bildung aus der DNA herausgeschnitten wird).

Die Expressionslevel der DNA-Methyltransferasen DNMT1 und DNMT3a weisen einen entgegengesetzten Trend auf: So nimmt die Expression der für die Erhaltungsmethylierung zuständigen Methyltransferase DNMT1 zwischen p2 und p20 stark ab. Dies ist verständlich, da in der entwickelten Retina kaum noch Zellteilungen vorkommen. Dagegen steigt das Expressionslevel der *de-novo*-Methyltransferase im gleichen Zeitraum an, so dass mutmaßlich eine rege *de-novo*-Methylierung in der entwickelten Retina stattfindet. Offenbar führt das Zusammenspiel von *de-novo*-Methylierung, hmC-Bildung durch mC-Oxidation und aktiver Demethylierung über fC zwischen p2 und p20 zu einem *steady-state*-Zustand, bei dem sich das globale DNA-Methylierungslevel nicht ändert, sondern konstant $3,0 \pm 0,1\%$ beträgt.

4.6.3 Zusammenfassung und Diskussion

Wie in den oben beschriebenen Projekten gezeigt werden konnte, gibt es bei Säugetieren im Laufe des Lebens offenbar beträchtliche Schwankungen des DNA-Modifikationsgrades. Diese Schwankungen betreffen nicht nur Methylcytosin, sondern ganz besonders die Modifikationen Hydroxymethyl- und Formylcytosin. Bei diesen Basen beobachtet man eindeutige, altersabhängige Trends.

Das hmC-Level steigt im menschlichen Hirn und in den untersuchten Mäusegeweben ab der Geburt stark an (Abbildung 32, Abbildung 34 und Abbildung 36). Beim Menschen endet dieser Anstieg in der grauen Substanz des Gehirns ungefähr ab dem 60. Lebensjahr. Ab diesem Zeitpunkt steigt das hmC-

Level nicht weiter an, sondern bleibt auf einem gleichbleibend hohem Niveau von durchschnittlich $1,23 \pm 0,07\%$. Würde man die von uns gemessene Zeitreihe bei der Maus fortführen, so würde man vermutlich im murinen Cerebrum und in der Retina, eventuell sogar in der Niere eine ähnliche Stagnation des hmC-Levels auf hohem Niveau beobachten, wie sie auch von Szulwach *et al.* im Hippocampus der Maus noch vor pw6 (= p42) beobachtet wurde.^[269] Leider erlaubte die starke Geräteauslastung keine weitere Fortführung dieser Zeitreihe.

In allen untersuchten Mäusegeweben beobachtet man ab den ersten Lebenstagen einen starken Rückgang des Formylcytosin-Levels um bis zu 85% (Abbildung 35 und Abbildung 36). Dieser Rückgang setzt in der Retina und in der Niere schon deutlich vor dem 10. bzw. 14. Lebenstag ein, während sich das fC-Level im Cerebrum bis dahin nicht ändert. Hier findet die starke Abnahme des DNA-Formylierungsgrades erst zwischen p14 und p90 statt.

Bei Methylcytosin sind die beobachteten Trends weniger stark ausgeprägt. Im menschlichen Hirn kann man nur in der grauen Substanz des Cerebrums einen Anstieg des mC-Levels beobachten (Abbildung 33). Dieser beginnt schon im Fötalstadium (15. Schwangerschaftswoche) und endet etwa im zweiundzwanzigsten Lebensjahr. Er ist mit ca. 45% außerdem weniger stark ausgeprägt als der Anstieg des hmC-Levels. Möglicherweise ist dieser Anstieg des mC-Levels hauptsächlich auf die von Lister *et al.* im menschlichen *Lobus frontalis* beobachteten „nichtklassischen“ *de-novo*-Methylierungen von mCH-Methylierungspositionen in der DNA zurückzuführen.^[282] Bei der Maus haben wir im Cerebrum und in der Niere einen Anstieg des mC-Levels um ca. 70% zwischen p14 und p90 beobachtet (Abbildung 34). Auch Lister *et al.* haben von einem solchen Anstieg des Methylierungsgrades im murinen Cerebrum berichtet. Dieser basiert genau wie beim Menschen auf *de-novo*-mCH-Methylierungen, beginnt ungefähr bei p7 und endet ca. bei pw6 (= p42).^[282] In den von uns gemessenen murinen Retinaprobe ändert sich dagegen das mC-Level zwischen p2 und p21 überraschenderweise nicht (Abbildung 36). Eventuell setzt hier der in den anderen Geweben beobachtete Anstieg des mC-Levels erst nach p21 ein.

Die von uns gemessenen Nukleosidlevel mit den dabei beobachteten altersabhängigen Dynamiken passen, wie z.T. schon oben ausgeführt, sehr gut zu den gängigen Theorien über diese epigenetischen Modifikationen.

So scheint es in der Maus keinen offensichtlichen Zusammenhang zwischen dem mC- und hmC-Level zu geben, der die von uns gemessenen hmC-Level direkt mit einer DNA-Demethylierung in Verbindung bringt: Obwohl das hmC-Level in den Geweben stark ansteigt, steigt gleichzeitig auch der DNA-Methylierungsgrad (Abbildung 34). Außerdem beobachten wir im vollständig entwickelten menschlichen Hirn eine Stagnation des hmC-Levels auf hohem Niveau. Beide Beobachtungen sind in guter Übereinstimmung mit der weithin akzeptierten Theorie, dass es sich beim Großteil des in der

DNA gefundenen hmCs um relativ stabile Modifikationen mit langer Lebensdauer handelt.^[40, 278, 290] Über die genaue epigenetische Funktion von hmC ist viel spekuliert worden.^[36, 278, 384-385, 522] Sicher ist, dass hydroxymethylierte DNA-Abschnitte eher mit einer aktiven Transkription verbunden sind.^[218, 229, 281] Möglicherweise genügt eine TET-vermittelte Umwandlung von methylierten zu hydroxymethylierten DNA-Abschnitten, um eine stabile und langfristige Expression vormals stillgelegter Gene zu ermöglichen. Erst wenn diese Gene ein weiteres Mal stabil stillgelegt werden sollen, findet eine Weiteroxidation von hmC zu fC und evtl. caC statt. Diese Basen können dann von TDG aus der DNA herausgeschnitten werden, so dass eine Remethylierung der dann unmodifizierten Cytosinbasen möglich ist.^[27, 34] Ein experimenteller Nachweis dieser oxidativen „Re-“methylierung steht jedoch aus.

Bei fC beobachten wir besonders hohe Level in den frühen Entwicklungsstadien der untersuchten Organe und Gewebe. Hier kommt es vermutlich zu einer Vielzahl von epigenetischen Reprogrammierungen auf DNA-Ebene. Wahrscheinlich ist dabei die Demethylierungsaktivität deutlich erhöht. Es ist bekannt, dass eine solche aktive DNA-Demethylierung über die Oxidation von mC zu fC oder caC möglich ist. Als Substrat der Glykosylase TDG wird fC aus der DNA herausgeschnitten und durch unmodifiziertes Cytosin ersetzt.^[34] Dies könnte die hohen fC-Level in den frühen und die wesentlich geringeren fC-Level in den späteren Entwicklungsstadien der untersuchten Organe und Gewebe erklären. Beim Großteil des von uns detektierten fCs würde es sich dann um transiente DNA-Modifikationen handeln, die kurz nach ihrer Entstehung wieder aus der DNA herausgeschnitten werden würden. In diesem Fall wären die von uns gemessenen fC-Level ein direkter Hinweis auf das Ausmaß an Demethylierungsaktivität in den untersuchten Proben.

4.6.4 Anhang

Da die in diesem Kapitel beschriebenen Ergebnisse zeitgleich mit der Anfertigung dieser Arbeit publiziert worden sind, finden sich alle Daten zu den hier gezeigten Graphen und die entsprechenden experimentellen Einzelheiten in diesen Publikationen bzw. in deren Anhang. ^[115, 169]

4.7 “Age-Dependent Levels of 5-Methyl-, 5-Hydroxymethyl-, and 5-Formylcytosine in Human and Mouse Brain Tissues”

M. Wagner[#], J. Steinbacher[#], T. F. Kraus[#], S. Michalakis, B. Hackner, T. Pfaffeneder, A. Perera, M. Müller, A. Giese, H. A. Kretzschmar, T. Carell, *Angew. Chem. Int. Ed. Engl.* **2015**, *54*, 12511-12514.

Co-first authors

Diese Publikation enthält, mit Ausnahme der in der Retina durchgeführten Studien, die im vorherigen Abschnitt vorgestellten Ergebnisse. Es wurden außerdem weitere Zeitpunkte in der postnatalen Entwicklung der Maus untersucht und in diese Publikation mitaufgenommen. Daneben finden sich hier auch erstmals Daten zum Vorkommen von fC in menschlichen Hirngewebe und in Neuronen.

Die Experimente, welche die Grundlage für diese Publikation bilden, sind alle von mir durchgeführt worden. Daneben habe ich die Daten ausgewertet, interpretiert und grafisch dargestellt. Auch an der Abfassung des Manuskripts war ich beteiligt. Bei der Untersuchung der Mäusegewebe bin ich von Jessica Steinbacher und Arshan Perera unterstützt worden.

Age-Dependent Levels of 5-Methyl-, 5-Hydroxymethyl-, and 5-Formylcytosine in Human and Mouse Brain Tissues**

Mirko Wagner, Jessica Steinbacher, Theo F. J. Kraus, Stylianos Michalakakis, Benjamin Hackner, Toni Pfaffeneder, Arshan Perera, Markus Müller, Armin Giese, Hans A. Kretzschmar†, and Thomas Carell*

Abstract: The absolute levels of 5-hydroxymethylcytosine (hmC) and 5-methylcytosine (mC) in human brain tissues at various ages were determined. Additionally, absolute levels of 5-formylcytosine (fC) in adult individuals and cytosine modification levels in sorted neurons were quantified. These data were compared with age-related fC, hmC, and mC levels in mouse brain samples. For hmC, an initial steady increase is observed, which levels off with age to a final steady-state value of 1.2 % in human brain tissue. This level is nearly twice as high as in mouse cerebral cortex. In contrast, fC declines rapidly with age during early developmental stages, thus suggesting that while hmC is a stable epigenetic mark, fC is more likely an intermediate of active DNA demethylation during early brain development. The trends in global cytosine modification dynamics during the lifespan of an organism are conserved between humans and mice and show similar patterns in different organs.

In addition to the four canonical Watson and Crick bases and 5-methylcytosine (mC), it has recently been discovered that mammalian DNA also contains 5-hydroxymethylcytosine (hmC),^[1] 5-formylcytosine (fC),^[2] and 5-carboxycytosine^[2b,3]. The biological role of these cytosine modifications is presently a focus of intense research, to which end the amount and distributions of the new bases need to be analyzed as the first step. Particularly large amounts of the base hmC were found in DNA isolated from mouse brain tissues.^[1,4] In cerebral cortex, for example, the hmC level reaches up to 0.8 % with respect to all present cytosines.^[4,5] Additionally, an increase in hmC content was observed during postnatal development in mice.^[4,6] These findings directly lead to the question of how hmC levels in humans vary over a lifetime, since knowledge of

age-dependent changes in hmC and fC levels is essential in order to gain insight into potential functions of these newly found bases.^[7] So far, however, information about cytosine modification levels in human tissues is limited. Herein we report absolute cytosine modification levels in human brain tissues at fetal and different postnatal developmental stages. We compare these data with modification levels previously reported by us for adult individuals.^[8] Quantification was performed by using our reported LC-MS-based isotope dilution method, in which synthetic isotopically labelled mC, hmC, and fC nucleosides are used as internal standards (Figure 1).^[5c,8,9]

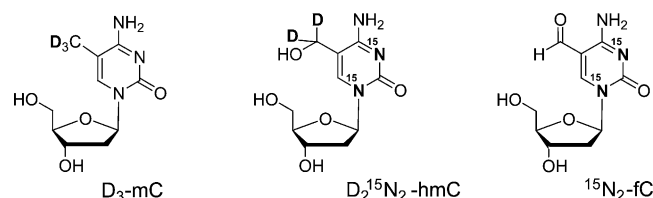


Figure 1. Depiction of the isotopically labeled mC, hmC, and fC nucleosides used as internal standards for mass spectrometry based quantification (D and ¹⁵N atoms are highlighted in bold).

For analysis, brain tissues from humans of different ages were provided by the BrainBank Munich. Tissues from a 15 week old fetus (15. WOP) and from four individuals between the ages of 0.6 and 88 years were analyzed. To this end, DNA was separately isolated from grey matter (cerebral cortex) and white matter of the cerebrum. The data obtained were compiled with previously reported data from four

[*] Dipl.-Chem. M. Wagner,^[‡] M. Sc. J. Steinbacher,^[‡] Dr. B. Hackner, Dr. T. Pfaffeneder, Dr. M. Müller, Prof. Dr. T. Carell
Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität München
Butenandtstr. 5-13, 81377 Munich (Germany)
E-mail: Thomas.Carell@lmu.de

Dr. T. F. J. Kraus,^[‡] Prof. Dr. A. Giese, Prof. Dr. H. A. Kretzschmar
Center for Neuropathology and Prion Research (ZNP)
Ludwig-Maximilians-Universität München
Feodor-Lynen-Str. 28, 81377 Munich (Germany)

Priv.-Doz. Dr. S. Michalakakis, M. Sc. A. Perera
Center for Integrated Protein Science at the Department of Pharmacy - Center for Drug Research
Ludwig-Maximilians-Universität München
Butenandtstr. 5-13, 81377 Munich (Germany)

[†] These authors contributed equally to this work.

[‡] deceased

[**] We thank the Deutsche Forschungsgemeinschaft (DFG) for financial support through the SFB program (SFB 749 and 1032). Further support was obtained from the Excellence Cluster (Center for Integrated Protein Science, CiPS^M). T.P. thanks the Fonds der Chemischen Industrie for a graduate fellowship.

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/anie.201502722>.

© 2015 The Authors. Published by Wiley-VCH Verlag GmbH & Co. KGaA. This is an open access article under the terms of the Creative Commons Attribution Non-Commercial NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

individuals aged 61 (two individuals), 84, and 85.^[8] In addition, from the 22 and 85 year old individuals, we were able to obtain DNA from cerebellum, again from both white and grey matter. Because prior studies with mouse tissues had revealed that cerebellum contains 40 % less hmC compared to cerebrum,^[4] a similar comparison in human tissues appeared highly desirable. Our quantification results regarding the absolute global levels of hmC in human brain tissues^[10] are summarized in Figure 2.

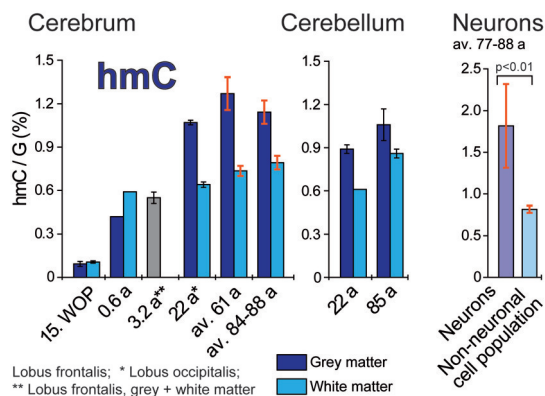


Figure 2. Age-related hmC levels in human cerebrum, cerebellum, and different brain cell populations. Values are given as modifications per 100 guanine bases. Guanine (G) was chosen as a reference because the overall G content is equal to the sum of cytosine (C) and the cytosine derivatives mC, hmC, and fC. In cerebrum and cerebellum, the white and grey matter were analyzed separately. Data for samples 61a (two individuals), 84a, and 85a (here for cerebrum only) were taken from Kraus et al.^[8] Cerebral occipital grey matter tissue of five human individuals aged 77–88 was sorted into neuronal and non-neuronal nuclei which were then analyzed separately. Black error bars represent the standard deviation of two to four technical replicates, thick red error bars represent the standard deviation of two to four biological replicates (the number of replicates was dependent on sample size). In the case of the 0.6 year old individual, owing to limited sample size, only one measurement could be performed. For statistical analysis, Student's unpaired two-tailed *t*-test was used. WOP=week of pregnancy, a=anno (year), av=average.

Our data show that in grey and white matter of the human cerebrum, the hmC level strongly increases with age. While in the cerebrum of a 15 week old fetus, only 0.1 % hmC is present in both white and grey matter, the fully developed human brain shows a constant hmC level of 1.2 % in the grey matter, which is rich in neuronal cell bodies.^[8] This is nearly twice the level detected in mouse cerebral cortex.^[4] In white matter, with myelinated axons surrounded by glial cells as the main components, hmC is less abundant. Nevertheless, the mean level of 0.75 % hmC observed in adults is higher than in most mouse brain tissues.^[4,8] Furthermore, in white matter the increase of hmC levels off already at the age of approximately one year. This is much earlier than in grey matter, where we observe a steady-state level only after the age of 22. Although white matter initially shows a higher hmC level compared to grey matter, this order is inverted during adolescence. Similar to cerebellum, the human cerebellum shows an age-dependent increase in hmC content, although at lower absolute levels. Our findings are in good agreement with mouse data available from others^[6] and our group.^[4]

We wondered whether the differences observed between the hmC levels in grey and white matter could be explained by large differences in the hmC content of neurons and glial cells. To answer this question, cerebral occipital cortex tissue samples from five individuals aged 77–88 were sorted into neuronal and non-neuronal nuclei. Cytosine modification content was then determined separately for both populations. This experiment indeed revealed a higher level of hmC in neurons compared to non-neuronal cells (1.82 % vs. 0.82 %; Figure 2).^[10d]

In order to relate our hmC data to cytosine methylation, we next quantified the levels of mC in human cerebrum and cerebellum at early developmental stages and during adolescence,^[6a, 10b–f, 11] again separately for grey and white matter. The resulting data, together with published data from four older individuals,^[8] are compiled in Figure 3.

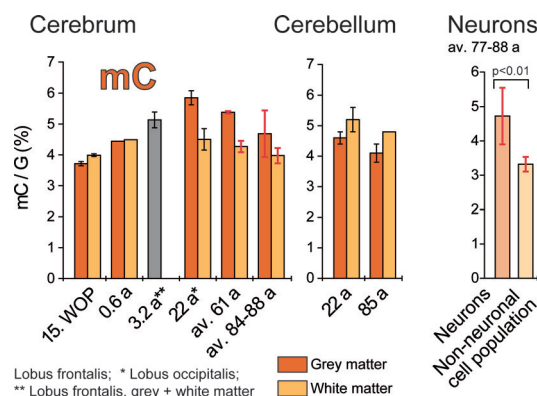


Figure 3. Age-related mC levels in human cerebrum, cerebellum, and different brain cell populations. Data for samples 61a (two individuals), 84a, and 85a (here for cerebrum only) were previously reported by us.^[8] Further details as in Figure 2.

In white matter, an age-dependent change for mC is barely detectable. Here, the mC level remains roughly constant at about 4.3 % during a human's lifetime. A moderate age effect is observed in the neuronal cell-body-rich grey matter, where mC values increase from 4.4 % (0.6a) to a maximum of roughly 6 % (22a). This finding is in full agreement with a recent study by Lister et al., who measured age-dependent *de novo* methylations in human and mouse lobus frontalis.^[6a]

In cerebellum, no age effect is observed. In the cell-sorting experiment, an elevated mC level was observed in neurons with respect to the non-neuronal cell population (Figure 3 and the Supporting Information).^[6a, 10d, 11a] In summary, our data show that global mC levels in humans change only slightly with age.

We next studied the levels of fC^[2a] in human brain tissues. This base^[2b] is proposed to be associated with active DNA demethylation.^[1b, 2, 3, 5a, 12] Because of limited human tissue availability, in combination with the fact that more tissue is needed to detect the low levels of fC,^[5c] data could only be obtained for the individuals aged 61 (sample 61*) and 85 (Figure 4).

In the cerebrum of adult human individuals, we observe fC levels of roughly $2\text{--}3 \times 10^{-4}$ % fC/G, which correspond to

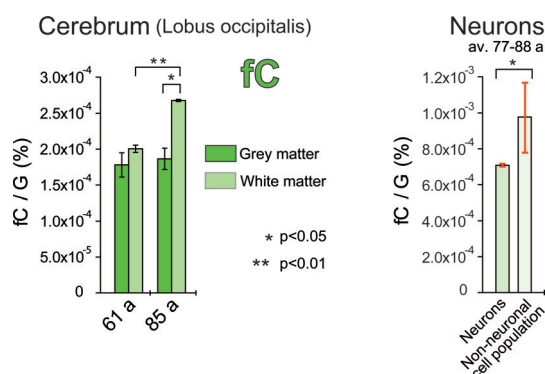


Figure 4. Global fC levels in cerebrum tissues from two human adults and in different brain cell populations. Further details as in Figure 2.

$4-8 \times 10^{-7}$ fC modifications per nucleoside (N). Similar levels of fC were found in the grey matter of the two individuals. By contrast, the fC levels differ in white matter. Here, the 85 year old individual showed an increased fC content in white compared to grey matter, while for the individual aged 61, comparable levels of fC were detected in both tissues. Consistent with this, the cell-sorting experiment revealed that with 9.8×10^{-4} % fC/G, more fC is present in the non-neuronal cell population, while in neurons, we observed a fC level of only 7.1×10^{-4} % (see Figure 4 and the Supporting Information). It should be noted that the fC levels observed by us in human cerebrum are as high as the level of fC in adult mouse cerebrum (see also Figure 5).^[5c] For a comparison of fC and hmC levels reported in this study with previously published data regarding the levels of these bases in mouse embryonic stem cells and organs at postnatal day 90, see Figures S1 and S2 in the Supporting Information.

To further study the age-dependent changes of cytosine modification levels in the mammalian brain and to overcome the limited human tissue availability, we expanded our investigations to mouse tissues. We analyzed fC, hmC,^[4,6,13] and mC^[6a,g,11c] levels in cerebrum tissue (cerebral cortex) sampled at postnatal day 1 (p1, newborn) and day 14 (p14), and from adult mice at postnatal day 90 (p90), postnatal month 12 (pm12), and postnatal month 18 (pm18; Figure 5).

Besides the expected increase in hmC content with age, we surprisingly observe a strong decrease in the fC modification level at early developmental stages.^[9b] Because fC is proposed to be generated from hmC by Tet-induced oxidation,^[2b] the decreasing fC level does not fit with the increasing amounts of hmC during brain development. One possible explanation for this counterintuitive observation is that fC is indeed an intermediate of an active DNA demethylation process^[11b,2,3,5a,12] and that active demethylation events in the brain cease with age. The higher fC levels in early lifetime suggest a faster hmC turnover via oxidation to fC,^[2b] followed by active demethylation.^[2,3,5a,12d,e] With increasing age, these as yet ill-defined processes may slow down, which would lead to the observed accumulation of hmC.

We observe similar trends in mouse kidney tissue, which was previously found to contain intermediate levels of hmC.^[5a] Here too, a low hmC level in young animals

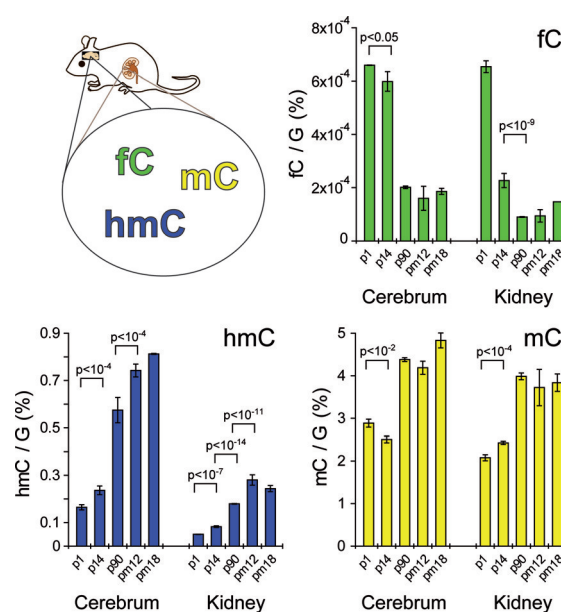


Figure 5. Global levels of fC, hmC, and mC in mouse cerebrum (cerebral cortex) and kidney at postnatal days 1 (p1) and 14 (p14) and in adult mice aged 90 days, 12 months, and 18 months. Values are given as modifications per 100 guanine bases. Guanine (G) was chosen as reference because it amounts to the sum of cytosine (C) and its derivatives mC, hmC, and fC. Samples were taken from wild-type mice (C57-BL6). Error bars represent the standard deviation of two (p1) or three (p14–pm18) biological replicates. For statistical analysis, Student's unpaired two-tailed *t*-test was used. pm = postnatal month.

corresponds to a high fC level, while higher hmC levels during later developmental stages go hand in hand with low fC values. This observation is in good agreement with the idea that reduced active demethylation is a reason for hmC accumulation in tissues during lifetime.

Finally, the mC level observed in mouse cerebrum (cerebral cortex) remains roughly constant until p14. It then shows a much stronger increase with age compared to human cerebrum.

In summary, we provide the first data on age-dependent variations in global hmC levels in humans over the entire lifespan. After a strong increase during early postnatal stages and adolescence, hmC reaches a steady-state level of 1.2 % in the fully developed brain.^[8] This is nearly twice as high as in mouse cerebral cortex.^[4] Furthermore, global levels of fC in brain tissues of human adults, and cytosine modification levels in human cerebral occipital cortex neurons were quantified. We also report age-dependent variations in fC, hmC, and mC levels in mouse tissues between postnatal day 1 and postnatal month 18. Our measurements revealed that the level of fC decreases strongly at early developmental stages, showing an inverted age-dependent trend compared to hmC. Our data support the idea that fC in the developing brain is mainly linked to active DNA demethylation,^[2,3,5a,12d,e] while hmC in this context is rather a stable epigenetic mark.^[14] The observed trends in global cytosine-modification dynamics during the lifespan of an organism are conserved^[15] between mammalian species^[16] and are similar in appearance in different organs.^[4,6,9b,10a,b,e,11,16,17]

Keywords: 5-formylcytosine · 5-hydroxymethylcytosine · aging · epigenetics · cerebral cortex

How to cite: *Angew. Chem. Int. Ed.* **2015**, *54*, 12511–12514
Angew. Chem. **2015**, *127*, 12691–12695

- [1] a) S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929–930; b) M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930–935.
- [2] a) T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeyer, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008–7012; *Angew. Chem.* **2011**, *123*, 7146–7150; b) S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300–1303.
- [3] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303–1307.
- [4] M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmler, S. Michalakakis, M. Müller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2010**, *49*, 5375–5377; *Angew. Chem.* **2010**, *122*, 5503–5505.
- [5] a) D. Globisch, M. Münzel, M. Müller, S. Michalakakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS ONE* **2010**, *5*, e15367; b) A. Szwagierczak, S. Bultmann, C. S. Schmidt, F. Spada, H. Leonhardt, *Nucleic Acids Res.* **2010**, *38*, e181; c) T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarchova, D. Schürmann, S. Michalakakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schär, M. Müller, T. Carell, *Nat. Chem. Biol.* **2014**, *10*, 574–581.
- [6] a) R. Lister, E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao, M. Esteller, C. He, F. G. Haghighi, T. J. Sejnowski, M. M. Behrens, J. R. Ecker, *Science* **2013**, *341*, 1237905; b) C.-X. Song, K. E. Szulwach, Y. Fu, Q. Dai, C. Yi, X. Li, Y. Li, C.-H. Chen, W. Zhang, X. Jian, J. Wang, L. Zhang, T. J. Looney, B. Zhang, L. A. Godley, L. M. Hicks, B. T. Lahn, P. Jin, C. He, *Nat. Biotechnol.* **2011**, *29*, 68–72; c) K. E. Szulwach, X. Li, Y. Li, C.-X. Song, H. Wu, Q. Dai, H. Irier, A. K. Upadhyay, M. Gearing, A. I. Levey, A. Vasanthakumar, L. A. Godley, Q. Chang, X. Cheng, C. He, P. Jin, *Nat. Neurosci.* **2011**, *14*, 1607–1616; d) L. Chouliaras, D. L. A. van den Hove, G. Kenis, S. Keitel, P. R. Hof, J. van Os, H. W. M. Steinbusch, C. Schmitz, B. P. F. Rutten, *Curr. Alzheimer Res.* **2012**, *9*, 536–544; e) H. Chen, S. Dzitoyeva, H. Manev, *Restor. Neurol. Neurosci.* **2012**, *30*, 237–245; f) M. A. Hahn, R. Qiu, X. Wu, A. X. Li, H. Zhang, J. Wang, J. Jui, S. G. Jin, Y. Jiang, G. P. Pfeifer, Q. Lu, *Cell Rep.* **2013**, *3*, 291–300.
- [7] a) W. Sun, L. Zang, Q. Shu, X. Li, *Genomics* **2014**, *104*, 347–351; b) M. Santiago, C. Antunes, M. Guedes, N. Sousa, C. J. Marques, *Genomics* **2014**, *104*, 334–340; c) L. Wen, F. Tang, *Genomics* **2014**, *104*, 341–346; d) D. L. A. van den Hove, L. Chouliaras, B. P. F. Rutten, *Curr. Alzheimer Res.* **2012**, *9*, 545–549.
- [8] T. F. J. Kraus, D. Globisch, M. Wagner, S. Eigenbrod, D. Widmann, M. Münzel, M. Müller, T. Pfaffeneder, B. Hackner, W. Feiden, U. Schüller, T. Carell, H. A. Kretzschmar, *Int. J. Cancer* **2012**, *131*, 1577–1590.
- [9] a) S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Hofner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593–14599; b) A. Perera, D. Eisen, M. Wagner, S. K. Laube, A. F. Künzel, S. Koch, J. Steinbacher, E. Schulze, V. Splith, N. Mittermeier, M. Müller, M. Biel, T. Carell, S. Michalakakis, *Cell Rep.* **2015**, *11*, 283–294.
- [10] a) T. Wang, Q. Pan, L. Lin, K. E. Szulwach, C.-X. Song, C. He, H. Wu, S. T. Warren, P. Jin, R. Duan, X. Li, *Hum. Mol. Genet.* **2012**, *21*, 5500–5510; b) L. Wen, X. Li, L. Yan, Y. Tan, R. Li, Y. Zhao, Y. Wang, J. Xie, Y. Zhang, C. Song, M. Yu, X. Liu, P. Zhu, X. Li, Y. Hou, H. Guo, X. Wu, C. He, R. Li, F. Tang, J. Qiao, *Genome Biol.* **2014**, *15*, R49; c) S.-G. Jin, X. Wu, A. X. Li, G. P. Pfeifer, *Nucleic Acids Res.* **2011**, *39*, 5015–5024; d) N. Coppieters, B. V. Dieriks, C. Lill, R. L. Faull, M. A. Curtis, M. Dragunow, *Neurobiol. Aging* **2014**, *35*, 1334–1344; e) T. F. Kraus, V. Guibourt, H. A. Kretzschmar, *J. Neural Transm.* **2014**, DOI: 10.1007/s00702-00014-01346-00704; f) S. F. Field, D. Beraldi, M. Bachman, S. K. Stewart, S. Beck, S. Balasubramanian, *PLoS ONE* **2015**, *10*, e0118202.
- [11] a) K. D. Siegmund, C. M. Connor, M. Campan, T. I. Long, D. J. Weisenberger, D. Biniszkiwicz, R. Jaenisch, P. W. Laird, S. Akbarian, *PLoS ONE* **2007**, *2*, e895; b) D. G. Hernandez, M. A. Nalls, J. R. Gibbs, S. Arepalli, M. van der Brug, S. Chong, M. Moore, D. L. Longo, M. R. Cookson, B. J. Traynor, A. B. Singleton, *Hum. Mol. Genet.* **2011**, *20*, 1164–1172; c) L. Chouliaras, D. L. A. van den Hove, G. Kenis, S. Keitel, P. R. Hof, J. van Os, H. W. M. Steinbusch, C. Schmitz, B. P. F. Rutten, *Neurobiol. Aging* **2012**, *33*, 1672–1681.
- [12] a) S. C. Wu, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 607–620; b) S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, *Nature* **2010**, *466*, 1129–1133; c) J. Guo, Y. Su, C. Zhong, G. Ming, H. Song, *Cell* **2011**, *145*, 423–434; d) A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334–35338; e) S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, A. Bellacosa, *Cell* **2011**, *146*, 67–79.
- [13] S. Dzitoyeva, H. Chen, H. Manev, *Neurobiol. Aging* **2012**, *33*, 2881–2891.
- [14] a) M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 1049–1055; b) M. A. Hahn, P. E. Szabo, G. P. Pfeifer, *Genomics* **2014**, *104*, 314–323.
- [15] R. D. Almeida, V. Sottile, M. Loose, P. A. De Sousa, A. D. Johnson, A. Ruzov, *Epigenetics* **2012**, *7*, 137–140.
- [16] T. Zheng, Q. Lv, X. Lei, X. Yin, B. Zhang, *Neurochem. Res.* **2015**, *40*, 688–697.
- [17] S. A. Tammen, G. G. Dolnikowski, L. M. Ausman, Z. Liu, K.-c. Kim, S. Friso, S.-W. Choi, *J. Cancer Prev.* **2014**, *19*, 301–308.

Received: March 24, 2015

Revised: May 13, 2015

Published online: July 3, 2015

4.8 “Low Values of 5-Hydroxymethylcytosine (5hmC), the “Sixth Base”, are Associated with Anaplasia in Human Brain Tumors”

T. F. J. Kraus, D. Globisch, **M. Wagner**, S. Eigenbrod, D. Widmann, M. Münzel, M. Müller, T. Pfaffeneder, B. Hackner, W. Feiden, U. Schüller, T. Carell, H. A. Kretzschmar, *Int. J. Cancer* **2012**, *131*, 1577-1590.

Ein typisches Charakteristikum von Tumorzellen und Tumorgeweben ist ihr aberrantes DNA-Methylierungsmuster. Nach der (Wieder)entdeckung der enzymatisch generierten DNA-Base hmC stellte sich daher sofort die Frage, ob in Tumorgeweben auch die DNA-Hydroxymethylierung dereguliert ist. Ziel dieses Projektes war es daher, in Zusammenarbeit mit der Arbeitsgruppe Kretzschmar vom Zentrum für Neuropathologie und Prionforschung (ZNP) diese Frage in Bezug auf Hirntumore zu beantworten. Dazu wurde uns eine ganze Reihe von Gewebeproben verschiedener humaner Hirntumortypen zur Verfügung gestellt. Mit der von uns entwickelten LC-MS-basierten Methode zur Quantifizierung der absoluten Level von modifizierten DNA-Nukleosiden sollten diese Proben analysiert werden.

In der hier vorliegenden Arbeit konnten wir zeigen, dass Tumorgewebe tatsächlich auch eine aberrante DNA-Hydroxymethylierung aufweisen. Im Gegensatz zu mC, wo man sowohl hypo- wie auch hypermethylierte Tumorzellen findet, wiesen alle von uns untersuchten Hirntumore eine ausgeprägte DNA-Hypohydroxymethylierung im Vergleich zu gesundem Hirngewebe auf. Dabei korreliert das detektierte hmC-Level negativ mit dem WHO-Tumorgrad und der Anaplasie, d.h. der malignen Entdifferenzierung der Tumorzellen. Das hmC-Level bietet sich deshalb als neuer Biomarker von malignen Zellveränderungen an.

In Zusammenarbeit mit Daniel Globisch und später auch mit Toni Pfaffeneder und Benjamin Hackner habe ich bei diesem Projekt den enzymatischen DNA-Verdau, die LC-MS-Messungen und die Datenanalyse und –interpretation durchgeführt. Daneben wurde auch die DNA aus den Tumorgeweben von mir isoliert.

Low values of 5-hydroxymethylcytosine (5hmC), the “sixth base,” are associated with anaplasia in human brain tumors

Theo F.J. Kraus^{1*}, Daniel Globisch^{2*}, Mirko Wagner², Sabina Eigenbrod¹, David Widmann¹, Martin Münzel², Markus Müller², Toni Pfaffeneder², Benjamin Hackner², Wolfgang Feiden³, Ulrich Schüller¹, Thomas Carell² and Hans A. Kretzschmar¹

¹ Center for Neuropathology and Prion Research (ZNP), Ludwig-Maximilians-University, Munich, Germany

² Department of Chemistry, Center for Integrated Protein Science (CIPS^M), Ludwig-Maximilians-University, Munich, Germany

³ Institute for Neuropathology, University Hospital Homburg, Homburg, Germany

5-Methylcytosine (5mC) in genomic DNA has important epigenetic functions in embryonic development and tumor biology. 5-Hydroxymethylcytosine (5hmC) is generated from 5mC by the action of the TET (Ten-Eleven-Translocation) enzymes and may be an intermediate to further oxidation and finally demethylation of 5mC. We have used immunohistochemistry (IHC) and isotope-based liquid chromatography mass spectrometry (LC-MS) to investigate the presence and distribution of 5hmC in human brain and brain tumors. In the normal adult brain, IHC identified 61.5% 5hmC positive cells in the cortex and 32.4% 5hmC in white matter (WM) areas. In tumors, positive staining of cells ranged from 1.1% in glioblastomas (GBMs) (WHO Grade IV) to 8.9% in Grade I gliomas (pilocytic astrocytomas). In the normal adult human brain, LC-MS also showed highest values in cortical areas (1.17% 5hmC/dG [deoxyguanosine]), in the cerebral WM we measured around 0.70% 5hmC/dG. 5hmC levels were related to tumor differentiation, ranging from lowest values of 0.078% 5hmC/dG in GBMs (WHO Grade IV) to 0.24% 5hmC/dG in WHO Grade II diffuse astrocytomas. 5hmC measurements were unrelated to 5mC values. We find that the number of 5hmC positive cells and the amount of 5hmC/dG in the genome that has been proposed to be related to pluripotency and lineage commitment in embryonic stem cells is also associated with brain tumor differentiation and anaplasia.

Epigenetics, the study of mechanisms that control gene expression (in a potentially heritable way), may be the most rapidly expanding field in tumor biology. On a molecular level, (i) DNA methylation, *i.e.*, covalent modification of cytosine bases resulting in 5-methylcytosine (5mC), (ii) histone modifications and (iii) nucleosome positioning are regarded as the driving epigenetic mechanisms. They are fundamental to the regulation of many cellular processes, including gene and micro

RNA expression, DNA-protein interactions, suppression of transposable element mobility, cellular differentiation, embryogenesis, X-chromosome inactivation and genomic imprinting.

In tumor biology, DNA methylation is the best-studied epigenetic change. Epigenetic silencing of O⁶-methylguanine DNA-methyltransferase (MGMT)¹ has been described as a strong predictive factor of treatment response to chemotherapy with alkylating agents of glioblastoma (GBM) and anaplastic astrocytoma (AA).² Methylation of CpG islands in the MGMT promoter with ensuing repression of MGMT transcriptional activity is generally viewed as the cause for this correlation.³

Oxidation of 5mC leading to 5-hydroxymethylcytosine (5hmC) has been identified as a new epigenetic phenomenon in mouse Purkinje cells.⁴ Three possible modes of action of 5hmC were discussed. (i) It might influence chromatin structure and local transcriptional activity by recruiting selective 5hmC-binding proteins or excluding 5mC-binding proteins. (ii) Conversion of 5mC to 5hmC might facilitate passive DNA demethylation by excluding the maintenance DNMT1, which recognizes 5hmC poorly. (iii) 5hmC may be an intermediate in a pathway of active DNA demethylation either by conversion to cytosine under certain conditions or by replacement of 5hmC by specific DNA repair mechanisms. Most recent findings support the hypothesis that 5hmC is an intermediate in a pathway for active DNA demethylation.^{5–7} The TET proteins, identified as a new family of enzymes that

Key words: epigenetics, brain tumors, 5-hydroxymethylcytosine, 5hmC, glioma, ependymoma, meningioma, IDH, TET

Grant sponsor: DFG Normalverfahren; **Grant numbers:** CA275/8-4, SFB 749, SFB 646; **Grant sponsor:** Fonds der Chemischen Industrie

DOI: 10.1002/ijc.27429

History: Received 18 Oct 2011; Accepted 29 Dec 2011; Online 10 Jan 2012

*T.F.J.K. and D.G. contributed equally to this work.

Correspondence to: Hans A. Kretzschmar, Center for Neuropathology and Prion Research (ZNP), Ludwig-Maximilians-University, 81377 Munich, Germany, E-mail: hans.kretzschmar@med.uni-muenchen.de or Thomas Carell, Center for Integrated Protein Science (CIPS^M), Department of Chemistry, Ludwig-Maximilians-University, 81377 Munich, Germany, E-mail: thomas.carell@cup.uni-muenchen.de

Table 1. Twenty-two tissue samples were selected from nine control cases from the BrainBank Munich

Case	Sample	Age	Gender	Postmortem time (h)	Region	Cause of death	5hmC positive cells (%)	5mC/dG (%)	5hmC/dG (%)
1	a	85	F	20	Occipital cortex	Cardiac arrest	55.05	5.16	1.17
1	b	85	F	20	Occipital white matter	Cardiac arrest	19.07	4.26	0.76
1	c	85	F	20	Frontal cortex	Cardiac arrest	70.29	–	–
1	d	85	F	20	Frontal white matter	Cardiac arrest	23.27	–	–
2	a	61	M	24	Frontal cortex	Cardiac arrest	55.46	5.05	1.06
2	b	61	M	24	Frontal white matter	Cardiac arrest	16.45	–	–
3	a	87	M	48	Frontal cortex	Cardiac arrest	54.32	5.94	1.13
3	b	87	M	48	Frontal white matter	Cardiac arrest	22.85	–	–
3	c	87	M	48	Occipital cortex	Cardiac arrest	54.70	5.64	1.14
3	d	87	M	48	Occipital white matter	Cardiac arrest	23.08	4.23	0.64
4	a	46	M	n.a.	Occipital cortex	Cardiac arrest	64.29	5.74	1.12
4	b	46	M	n.a.	Occipital white matter	Cardiac arrest	36.34	4.36	0.54
5	a	84	M	19	Frontal white matter	Cardiac arrest	28.57	3.82	0.78
5	b	84	M	19	Frontal cortex	Cardiac arrest	59.92	4.55	1.19
6	a	61	M	n.a.	Frontal white matter	Cardiac arrest	61.61	4.14	0.76
6	b	61	M	n.a.	Frontal cortex	Cardiac arrest	70.33	5.35	1.35
7	a	26	F	n.a.	Frontal cortex	Hemorrhagic pericard effusion	80.17	–	–
7	b	26	F	n.a.	Frontal white matter	Hemorrhagic pericard effusion	58.06	–	–
8	a	41	M	n.a.	Frontal cortex	Multiple organ failure	65.02	–	–
8	b	41	M	n.a.	Frontal white matter	Multiple organ failure	38.70	–	–
9	a	39	M	n.a.	Frontal cortex	Haematothorax	46.75	–	–
9	b	39	M	n.a.	Frontal white matter	Haematothorax	30.00	–	–

alter the methylation status of DNA, are 2-oxoglutarate (2OG)- and Fe(II)-dependent enzymes that catalyze the conversion of 5mC to 5hmC,⁸ 5-formylcytosine and 5-carboxylcytosine.^{5–7} TET proteins and 5hmC have been reported in various tissues and both are tightly regulated during embryonic stem cells (ESC) differentiation.⁹ 5hmC content has been described as significantly reduced in stem cell/progenitor cell compartments and in human cancers (prostate, breast and colon).¹⁰

To obtain a first insight into a possible role of 5hmC in the human brain, we used isotope-based liquid chromatography mass spectrometry (LC-MS) to measure 5hmC concentrations in various cortical and white matter regions of the adult human brain to elucidate whether hydroxymethylation may be present in human brain tumors, we measured 5hmC content in slowly growing frequent brain tumors, *i.e.*, astrocytomas WHO Grades I, II and rapidly growing AAs WHO Grade III and GBMs (WHO Grade IV)¹¹ and other brain tumors. In addition, we used immunohistochemistry (IHC) to investigate the distribution of 5hmC in the brain and in brain tumor cells.

Material and Methods

Sample selection normal brain regions

To evaluate the amount of 5hmC in human brain, we selected tissue of nine donors that had provided their brains

for research purposes after death. The samples were obtained from the Brain Bank Munich. As target regions we selected frontal cortex, occipital cortex, frontal white matter and occipital white matter. The tissue samples had either been formalin-fixed and paraffin-embedded (FFPE) or stored at -80°C . For this study, we chose 22 human control samples, well characterized by age and gender (Table 1). The age of the patients varied from 26 to 87 with a mean age of 59 years.

Sample selection tumors

To evaluate 5hmC in human brain tumors, we used FFPE and frozen material that had been stored at -80°C . The samples were obtained from the Brain Tumor Bank, Center for Neuropathology, LMU Munich, as well as from the Brain Tumor Bank of the Institute for Neuropathology, University Hospital Homburg. For this study, we selected areas consisting of more than 75% of tumor cells as estimated from freshly made and H&E stained sections. In each case we separated clearly visible tumor areas. One hundred seventeen human tumor samples derived from 117 different tumors were selected, including 80 astrocytomas and GBMs, 23 ependymomas and 14 meningiomas (Table 2). The age of the patients ranged from 3 to 81 years with a mean of 46 years and the male to female ratio was 67:43. In the astrocytoma

Table 2. One hundred seventeen tumor samples were selected from the Brain Tumour Bank, Center for Neuropathology, LMU Munich and the Brain Tumour Bank of the Institute for Neuropathology, University Hospital Homburg

Case	Age	Gender	Tumor	Grade	Region	5hmC positive cells (%)	5mC/dG (%)	5hmC/dG (%)	IDH status
1	15	M	Pilocytic astrocytoma (WHO Grade I)	I	Temporal	13.66	4.75	0.18	No mutation
2	23	F	Pilocytic astrocytoma (WHO Grade I)	I	Thoracic spine	10.83	4.10	0.24	No mutation
3	23	M	Pilocytic astrocytoma (WHO Grade I)	I	n.a.	7.69	4.58	0.24	No mutation
4	12	F	Pilocytic astrocytoma (WHO Grade I)	I	Rhombencephalon	10.03	4.99	0.17	No mutation
5	9	M	Pilocytic astrocytoma (WHO Grade I)	I	Cerebellum	1.66	4.98	0.12	No mutation
6	21	M	Pilocytic astrocytoma (WHO Grade I)	I	Temporal	9.57	–	–	No mutation
7	34	M	Diffuse astrocytoma (WHO Grade II)	II	Fronto-temporal	3.75	4.66	0.17	Mutation
8	35	M	Diffuse astrocytoma (WHO Grade II)	II	Insula	1.47	–	–	Mutation
9	36	F	Diffuse astrocytoma (WHO Grade II)	II	Occipital	0.72	–	–	Mutation
10	15	M	Diffuse astrocytoma (WHO Grade II)	II	Thalamus	3.35	–	–	No mutation
11	28	F	Diffuse astrocytoma (WHO Grade II)	II	Precentral	1.55	–	–	Mutation
12	n.a.	n.a.	Diffuse astrocytoma (WHO Grade II)	II	n.a.	–	4.50	0.09	Mutation
13	n.a.	n.a.	Diffuse astrocytoma (WHO Grade II)	II	n.a.	–	4.57	0.24	Mutation
14	44	M	Diffuse astrocytoma (WHO Grade II)	II	Insula	7.04	–	–	No mutation
15	66	F	Diffuse astrocytoma (WHO Grade II)	II	Insula	5.19	–	–	No mutation
16	35	M	Diffuse astrocytoma (WHO Grade II)	II	Pons	9.06	–	–	No mutation
17	41	F	Diffuse astrocytoma (WHO Grade II)	II	Tectum	11.89	–	–	No mutation
18	30	M	Diffuse astrocytoma (WHO Grade II)	II	Temporal	0.00	–	–	Mutation
19	3	M	Diffuse astrocytoma (WHO Grade II)	II	Basal Ganglia	–	5.53	0.46	No mutation
20	74	M	Anaplastic astrocytoma (WHO Grade III)	III	Fronto-temporal	0.59	4.40	0.08	Mutation
21	43	M	Anaplastic astrocytoma (WHO Grade III)	III	Frontal	–	4.34	0.08	Mutation
22	41	F	Anaplastic astrocytoma (WHO Grade III)	III	Parietal	–	4.51	0.10	Mutation
23	59	M	Anaplastic astrocytoma (WHO Grade III)	III	n.a.	7.27	–	–	No mutation
24	62	M	Anaplastic astrocytoma (WHO Grade III)	III	Central	3.72	–	–	Mutation
25	41	M	Anaplastic astrocytoma (WHO Grade III)	III	Frontal	2.40	–	–	Mutation
26	37	F	Anaplastic astrocytoma (WHO Grade III)	III	Frontal	2.29	–	–	Mutation
27	37	F	Anaplastic astrocytoma (WHO Grade III)	III	Temporal	2.68	–	–	No mutation
28	31	M	Anaplastic astrocytoma (WHO Grade III)	III	Temporal	0.80	–	–	Mutation
29	32	M	Anaplastic astrocytoma (WHO Grade III)	III	Temporal	2.69	–	–	Mutation
30	42	M	Anaplastic astrocytoma (WHO Grade III)	III	Temporal	0.13	–	–	Mutation
31	44	F	Anaplastic astrocytoma (WHO Grade III)	III	n.a.	6.85	–	–	No mutation
32	44	F	Anaplastic astrocytoma (WHO Grade III)	III	Frontal	4.86	–	–	No mutation
33	81	F	Anaplastic astrocytoma (WHO Grade III)	III	Cerebellum	1.11	–	–	No mutation
34	70	M	Anaplastic astrocytoma (WHO Grade III)	III	Temporal	5.35	–	–	No mutation
35	61	M	Anaplastic astrocytoma (WHO Grade III)	III	Thalamus	6.00	–	–	No mutation
36	53	M	Anaplastic astrocytoma (WHO Grade III)	III	Temporo-mesial	–	4.76	0.30	No mutation
37	50	M	Anaplastic astrocytoma (WHO Grade III)	III	Temporal	–	3.79	0.10	No mutation
38	72	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporoparietal	0.37	4.35	0.16	No mutation
39	47	M	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	–	4.33	0.13	No mutation
40	61	M	Glioblastoma multiforme (WHO Grade IV)	IV	Parietal	2.26	4.06	0.10	No mutation
41	69	F	Glioblastoma multiforme (WHO Grade IV)	IV	Occipital	0.25	3.77	0.09	No mutation

Table 2. One hundred seventeen tumor samples were selected from the Brain Tumour Bank, Center for Neuropathology, LMU Munich and the Brain Tumour Bank of the Institute for Neuropathology, University Hospital Homburg (Continued)

Case	Age	Gender	Tumor	Grade	Region	5hmC positive cells (%)	5mC/dG (%)	5hmC/dG (%)	IDH status
42	55	M	Glioblastoma multiforme (WHO Grade IV)	IV	Parietal	0.08	5.58	0.03	No mutation
43	44	F	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	0.28	4.70	0.09	No mutation
44	64	M	Glioblastoma multiforme (WHO Grade IV)	IV	Parietal	0.97	4.32	0.07	No mutation
45	25	F	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	1.37	4.70	0.03	Mutation
46	66	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	3.48	4.27	0.05	No mutation
47	64	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.49	4.06	0.11	No mutation
48	67	F	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.00	–	–	No mutation
49	67	M	Glioblastoma multiforme (WHO Grade IV)	IV	Occipital	0.55	–	–	No mutation
50	41	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.67	–	–	No mutation
51	66	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	1.82	–	–	No mutation
52	58	M	Glioblastoma multiforme (WHO Grade IV)	IV	Parietal	1.82	–	–	No mutation
53	30	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	1.22	–	–	Mutation
54	69	F	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	1.45	–	–	No mutation
55	53	f	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	1.60	–	–	No mutation
56	43	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporo-parieto-occipital	2.21	–	–	No mutation
57	59	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	1.14	–	–	No mutation
58	7	M	Glioblastoma multiforme (WHO Grade IV)	IV	Thalamus	2.28	–	–	No mutation
59	67	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.78	–	–	No mutation
60	64	F	Glioblastoma multiforme (WHO Grade IV)	IV	Temporo-mesial	1.61	–	–	No mutation
61	43	F	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	1.19	–	–	Mutation
62	72	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	1.49	–	–	No mutation
63	32	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.53	–	–	Mutation
64	33	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	1.20	–	–	Mutation
65	46	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	1.92	–	–	Mutation
66	40	F	Glioblastoma multiforme (WHO Grade IV)	IV	Parieto-occipital	0.96	–	–	Mutation
67	38	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.60	–	–	Mutation
68	61	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.31	–	–	No mutation
69	47	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.62	–	–	No mutation
70	78	M	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.38	–	–	No mutation
71	71	F	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	1.23	–	–	No mutation
72	62	F	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	0.43	–	–	No mutation
73	64	F	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	1.89	–	–	No mutation
74	70	M	Glioblastoma multiforme (WHO Grade IV)	IV	Parietal	0.87	–	–	No mutation
75	41	F	Glioblastoma multiforme (WHO Grade IV)	IV	Parietal	0.40	–	–	No mutation
76	74	M	Glioblastoma multiforme (WHO Grade IV)	IV	Postcentral	0.53	–	–	No mutation
77	55	M	Glioblastoma multiforme (WHO Grade IV)	IV	Thalamus	0.60	–	–	No mutation
78	37	F	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	–	4.36	0.08	Mutation
79	34	M	Glioblastoma multiforme (WHO Grade IV)	IV	Frontal	–	4.03	0.04	Mutation
80	34	F	Glioblastoma multiforme (WHO Grade IV)	IV	Temporal	–	3.68	0.04	Mutation
81	30	F	Myxopapillary ependymoma (WHO Grade I)	I	Lumbal spine	44.14	5.19	0.09	n.a.
82	68	M	Subependymoma (WHO Grade I)	I	Rhombencephalon	36.50	4.57	0.42	n.a.

Table 2. One hundred seventeen tumor samples were selected from the Brain Tumour Bank, Center for Neuropathology, LMU Munich and the Brain Tumour Bank of the Institute for Neuropathology, University Hospital Homburg (Continued)

Case	Age	Gender	Tumor	Grade	Region	5hmC positive cells (%)	5mC/dG (%)	5hmC/dG (%)	IDH status
83	20	M	Subependymoma (WHO Grade I)	I	Lateral ventricle	13.54	–	–	n.a.
84	59	M	Subependymoma (WHO Grade I)	I	Lateral ventricle	28.95	–	–	n.a.
85	54	M	Subependymoma (WHO Grade I)	I	Rhombencephalon	16.49	4.55	0.35	n.a.
86	46	F	Subependymoma (WHO Grade I)	I	Thoracic spine	1.28	–	–	n.a.
87	46	F	Subependymoma (WHO Grade I)	I	Lateral ventricle	16.81	–	–	n.a.
88	46	M	Subependymoma (WHO Grade I)	I	Rhombencephalon	15.22	–	–	n.a.
89	n.a.	n.a.	Subependymoma (WHO Grade I)	I	n.a.	–	4.63	0.20	n.a.
90	n.a.	n.a.	Subependymoma (WHO Grade I)	I	n.a.	–	4.58	0.37	n.a.
91	30	F	Ependymoma (WHO Grade II)	II	Filum	20.09	5.04	0.09	n.a.
92	41	M	Ependymoma (WHO Grade II)	II	Cervical spine	5.15	4.74	0.17	n.a.
93	39	M	Ependymoma (WHO Grade II)	II	Cervical spine	6.26	3.38	0.25	n.a.
94	70	F	Ependymoma (WHO Grade II)	II	Cervical spine	15.85	3.70	0.41	n.a.
95	23	M	Ependymoma (WHO Grade II)	II	Thoracic spine	14.43	6.16	0.21	n.a.
96	54	M	Ependymoma (WHO Grade II)	II	Cervical spine	10.86	5.01	0.12	n.a.
97	42	M	Ependymoma (WHO Grade II)	II	n.a.	12.02	–	–	n.a.
98	76	F	Ependymoma (WHO Grade II)	II	Lumbal spine	15.04	–	–	n.a.
99	21	M	Cellular ependymoma (WHO Grade II)	II	Lateral ventricle	12.49	5.05	0.13	n.a.
100	4	M	Anaplastic ependymoma (WHO Grade III)	III	Rhombencephalon	24.16	4.95	0.23	n.a.
101	5	F	Anaplastic ependymoma (WHO Grade III)	III	Posterior cranial fossa	10.47	4.22	0.03	n.a.
102	59	F	Anaplastic ependymoma (WHO Grade III)	III	Thoracic spine	16.59	4.78	0.32	n.a.
103	6	F	Anaplastic ependymoma (WHO Grade III)	III	Lateral ventricle	1.25	–	–	n.a.
104	72	F	Transitional meningioma (WHO Grade I)	I	Fronto-basal	32.12	4.48	0.12	n.a.
105	22	F	Meningotheliomatous meningioma (WHO Grade I)	I	Frontal	29.06	4.04	0.17	n.a.
106	60	F	Transitional meningioma (WHO Grade I)	I	Cerebellum	9.52	4.08	0.20	n.a.
107	59	M	Fibrous meningioma (WHO Grade I)	I	Occipital	6.82	–	–	n.a.
108	28	M	Meningothelial meningioma (WHO Grade I)	I	Precentral	18.30	–	–	n.a.
109	79	F	Meningothelial meningioma (WHO Grade I)	I	Temporal	12.46	–	–	n.a.
110	61	M	Atypical meningioma (WHO Grade II)	II	Frontal	1.11	4.65	0.08	n.a.
111	33	F	Chordoid meningioma (WHO Grade II)	II	Frontal	5.24	–	–	n.a.
112	72	F	Atypical meningioma (WHO Grade II)	II	Frontal	14.34	–	–	n.a.
113	79	F	Atypical meningioma (WHO Grade II)	II	Frontal	2.93	4.44	0.11	n.a.
114	26	M	Atypical meningioma (WHO Grade II)	II	Skullbase	8.94	–	–	n.a.
115	n.a.	n.a.	Atypical meningioma (WHO Grade II)	II	n.a.	–	4.24	0.27	n.a.
116	n.a.	n.a.	Atypical meningioma (WHO Grade II)	II	n.a.	–	4.29	0.08	n.a.
117	n.a.	n.a.	Atypical meningioma (WHO Grade II)	II	n.a.	–	3.98	0.16	n.a.

and GBM group, the mean age was 47 years, in ependymomas 40 years and in meningiomas 54 years.

Immunohistochemistry and cell counting

FFPE tissue samples were used for IHC. Standard protocols were followed, in brief 4-μm FFPE sections were deparaffi-

nized by subsequent incubation in xylene (30 min), 100% ethanol (10 min), 96% ethanol (5 min) and 70% ethanol (5 min). The sections were then treated with 2N HCl (35°C, 20 min), pre-incubated with blocking solution (3% I-Block Protein-Based Blocking Reagent (Applied Biosystems, Darmstadt, Germany)) and incubated with the diluted primary

antibody for 18 hr at 4°C. As primary antibody, we used polyclonal rabbit anti-5hmC diluted 1:1,000 (Active Motif, Rixensart, Belgium). Subsequently, the sections were treated with Super Sensitive Polymer-Horseradish peroxidase (HRP) Detection System (BioGenix, San Ramon, CA), followed by incubation in 3,3'-diaminobenzidine (DAB) for 5 min and staining in hemalum. For immunofluorescence staining, we followed standard protocols. As primary antibody, we used polyclonal rabbit anti-5hmC diluted 1:1,000 (Active Motif, Rixensart, Belgium) and monoclonal mouse anti-Ki-67 diluted 1:500 (Dako, Hamburg, Germany). As secondary antibodies, we used Alexa Fluor 488 goat-anti-rabbit IgG diluted 1:300 (Invitrogen, Darmstadt, Germany) and Alexa Fluor 546 goat-anti-mouse IgG diluted 1:300 (Roth, Karlsruhe, Germany). For nuclear staining, we used 4',6-diamidin-2-phenylindol (Roth, Karlsruhe, Germany). Finally, we used Dako fluorescence mounting medium (Dako, Hamburg, Germany).

For cell counting, randomly selected sections were photographed at $\times 10$ and $\times 20$ magnification depending on tissue type and size. Following that at least 200 cells were counted in each case.

To identify IDH1 mutations in astrocytoma and GBM, FFPE samples were stained using the monoclonal mouse IDH1^{R132H} antibody (Dianova, Hamburg, Germany), diluted 1:20, according to the manufacturer's protocol on a Ventana BenchMark system (Roche, Basel, Switzerland).

DNA extraction

For LC-MS, DNA isolation was performed on the basis of the QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol with slight modifications. The RNA digestion was executed twice and instead of column purification phenol extraction was performed, i.e., after addition of buffer AL, the sample was mixed and incubated. The sample was distributed equally to two 2 mL reaction tubes, if necessary. A 1/1 mixture of Roti®Phenol/chloroform (1 vol) was added and the tube was shaken vigorously at RT for 5 min. The tube was centrifuged (12,100g, 15 min) and the aqueous layer was collected. This procedure was repeated once. To the obtained aqueous layer chloroform (1 vol) was added and the tube was shaken at RT for 1 min. After centrifugation (12,100g, 5 min) the aqueous layer was collected. During collection of the aqueous layer, special care was taken to include the interphase. The sample was distributed equally to two 2 mL reaction tubes, if necessary. Ethanol (3 vol) was added. The sample was left to stand at RT for approximately 2 h. After precipitation of the DNA the tube was centrifuged (12,100g, 30 min). The supernatant was discarded and the pellet was dried. Subsequently, it was dissolved in water (100–400 μ L). The solution was centrifuged (12,100g, 30 min) and the supernatant was collected.

Additionally, we isolated DNA from FFPE samples for IDH mutation analysis using a Maxwell 16 system (Promega, Madison, WI) and Maxwell 16 FEV FFPE kit (Promega, Madison, WI) according to the manufacturer's protocols.

Enzymatic digestion

For the enzymatic digestion DNA mixtures (4–10 μ g in a final volume of 100 μ L H₂O) were heated to 100°C for 5 min to denature the DNA and rapidly cooled on ice. After addition of standard solutions, buffer A (10 μ L, 300 mM ammonium acetate, 100 mM CaCl₂, 1 mM ZnSO₄, pH 5.7) and nuclease S1 (80 units, *Aspergillus oryzae*) the mixture was incubated for 3 hr at 37°C. Addition of buffer B (12 μ L, 500 mM Tris-HCl, 1 mM EDTA), antarctic phosphatase (10 units), snake venom phosphodiesterase I (0.2 units, *Crotalus adamanteus* venom) and incubation for further 3 hr at 37°C completed the digestion. Labeled nucleosides d₂-5hmC and d₃-5mC were added, followed by centrifugation of the sample (12,100g, 15 min). Each sample was performed at least in duplicate with independent concentrations of the two labeled nucleosides. The concentrations of standard solutions were chosen to be in the expected range of the sample nucleoside concentration.

LC-MS

The samples (100 μ L injection volume) were analyzed by LC-MS on a Thermo Finnigan LTQ Orbitrap XL and were chromatographed by a Dionex Ultimate 3000 HPLC system with a flow of 0.15 mL/min over an Uptisphere120-3HDO column from Interchim. The column temperature was maintained at 30°C. Eluting buffers were buffer C (2 mM HCOONH₄ in H₂O (pH 5.5)) and buffer D (2 mM HCOONH₄ in H₂O/MeCN 20/80 (pH 5.5)). The gradient was 0 \rightarrow 12 min; 0% \rightarrow 3% buffer D; 12 \rightarrow 60 min; 3% \rightarrow 60% buffer D; 60 \rightarrow 62 min; 60% \rightarrow 100% buffer D; 62 \rightarrow 70 min; 100% buffer D; 70 \rightarrow 85 min; 100 \rightarrow 0% buffer D; 85 \rightarrow 95 min; 0 % buffer D. The elution was monitored at 260 nm (Dionex Ultimate 3000 Diode Array Detector). The chromatographic eluent was directly injected into the ion source without prior splitting. Ions were scanned by the use of a positive polarity mode over a full-scan range of m/z 200–1,000 with a resolution of 30,000. Parameters of the mass spectrometer were tuned with a freshly mixed solution of adenosine (5 μ M) in buffer C. The parameters used in this section were sheath gas flow rate, 16 arb; auxiliary gas flow rate, 11 arb; sweep gas flow rate, 4 arb; spray voltage, 5.0 kV; capillary temperature, 200°C; capillary voltage, 12 V, tube lens 86 V.

Pyrosequencing of DNA

To confirm IDH1 mutations detected by IHC staining, we sequenced a 88 bp long fragment of the IDH1 gene including codon 132. The IDH1 primers were designed using PSQ Assay design software (Qiagen, Hilden, Germany). As forward primer, we used 5'-biotin-AAAAATATCCCCCG GCTTG-3' and as reverse primer 5'-TGCCAACATGACTTA CTTGATCC-3'. PCR reactions were performed using the HotStarTaq DNA polymerase master mix (Qiagen, Hilden, Germany) according to the manufacturer's protocol and a

standard PCR protocol. The PCR product was then sequenced using a PyroMark Q24 System (Qiagen, Hilden, Germany) and the Pyro Gold reagent kit (Qiagen, Hilden, Germany). We used 5'-TGATCCCCATAAGCAT-3' as sequencing primer. The subsequent processing was performed according to the manufacturer's protocols. Data analysis was performed using PyroMark Q24 software (Qiagen, Hilden, Germany).

Statistical analysis

Statistical calculations were performed using ANOVA single factor analysis and Tukey-Kramer *post hoc* test as well as unpaired *t*-test. Analysis was performed using GraphPad Prism 5.

Results

Not all cells express 5hmC in normal human brain and brain tumors at detectable levels

IHC with an antibody against 5hmC in general showed either strong staining or no staining with only a small number of faintly stained nuclei (Fig. 1). In normal human frontal brain, the majority of neurons of the cortex showed strong reactivity for 5hmC (Fig. 1a, upper right corner). In contrast, there was weaker immunohistochemical staining for 5hmC in the frontal subcortical white matter (Fig. 1a, lower left corner). About 70% of WM cells showed negative staining; these cells had mostly round, small nuclei and appeared to be oligodendrocytes. In the occipital cortex (Fig. 1b) and subcortical white matter (Fig. 1c), we found comparable results. In human brain tumors, we found significantly lower numbers of 5hmC positive cells. In astrocytomas and GBMs (Figs. 1d–1g), the number of 5hmC positive cells decreased from pilocytic astrocytomas (Fig. 1d) to diffuse astrocytomas (Fig. 1e), AAs (Fig. 1f) and GBMs with mean of 1.43% positive cells (Fig. 1g). Whilst there were about 10% of positive cells in pilocytic astrocytomas, GBMs showed only very few positive cells. In ependymomas (Figs. 1h–1j), we found that most of the cells were not stained. Grade I ependymomas (Fig. 1h) showed more 5hmC positive cells than Grade II (Fig. 1i) and Grade III tumors (Fig. 1j). Grades II and III ependymomas showed more or less equal staining but anaplastic ependymomas showed a weaker staining for 5hmC. Furthermore, there were strong regional variations. In meningiomas (Figs. 1k–1l), we found similar results to ependymomas. Grade I meningiomas (Fig. 1k) showed more 5hmC positive cells than Grade II tumors (Fig. 1l) and Grade II meningiomas showed a weaker staining for 5hmC with regional variations. We found that staining for 5hmC and the proliferation-associated Ki-67 was mutually exclusive; an example of this general finding is shown in Figures 1m–1o. We often observed a rhythmic variation of areas with higher proliferative activity (Ki-67) and sparse 5hmC staining and *vice versa*.

There are significantly more 5hmC positive stained cells in the cortex compared to white matter areas. 5hmC/dG in human brain is significantly higher in cortex than white matter

When we counted the 5hmC positive and negative cells the values for various cortical areas did not show significant differences (Fig. 2a, $p > 0.05$, ANOVA and Tukey-Kramer *post hoc* test). Furthermore, we did not find any significant difference in the ration of 5hmC positive cells in various white matter areas (Fig. 2a, $p > 0.05$, ANOVA and Tukey-Kramer *post hoc* test). However, we found significant differences comparing frontal cortex and frontal white matter areas (Fig. 2a, $p < 0.01$, ANOVA and Tukey-Kramer *post hoc* test) as well as occipital cortex and occipital white matter areas (Fig. 2a, $p < 0.05$, ANOVA and Tukey-Kramer *post hoc* test).

Additionally, we performed LC-MS measurements. 5hmC was found in measurable amounts in all brain regions but there was no significant difference between various cortical (Fig. 2b, $p > 0.05$, ANOVA and Tukey-Kramer *post hoc* test) and various white matter areas (Fig. 2b, $p > 0.05$, ANOVA and Tukey-Kramer *post hoc* test). But the values measured in the frontal cortex were significantly higher than in the frontal white matter (Fig. 2b, $p < 0.01$, ANOVA and Tukey-Kramer *post hoc* test) and the values measured in the occipital cortex were significantly higher than in the occipital white matter (Fig. 2b, $p < 0.001$, ANOVA and Tukey-Kramer *post hoc* test).

The 5mC values measured by LC-MS did not show significant differences between different cortical areas ($p > 0.05$ using ANOVA and Tukey-Kramer *post hoc* test, Fig. 2c) nor between different white matter areas ($p > 0.05$ using ANOVA and Tukey-Kramer *post hoc* test, Fig. 2c) but we found significant differences between cortical and white matter areas ($p < 0.05$, ANOVA and Tukey-Kramer *post hoc* test, Fig. 2c).

The number of 5hmC positive cells and 5hmC/dG levels in human brain tumors are related to WHO grade

One hundred seventeen human tumors were selected for analysis including 6 pilocytic astrocytomas (Grade I), 13 diffuse astrocytomas (Grade II), 18 AAs (Grade III) and 43 GBMs (Grade IV), 10 ependymomas Grade I, 9 ependymomas (Grade II) and 4 anaplastic ependymomas (Grade III), 6 WHO Grade I meningiomas and 8 atypical meningiomas (Grade II).

In WHO Grade I tumors, the mean of 5hmC positive cells was 16.73%, in Grade II 7.87%, in Grade III 5.51% and in Grade IV 1.07% (Fig. 2d) with a highly significant difference between Grades I and II ($p < 0.001$, ANOVA and Tukey-Kramer *post hoc* test), Grades I and III ($p < 0.001$, ANOVA and Tukey-Kramer *post hoc* test), Grades I and IV ($p < 0.001$, ANOVA and Tukey-Kramer *post hoc* test) as well as Grades II and IV ($p < 0.001$, ANOVA and Tukey-Kramer *post hoc* test).

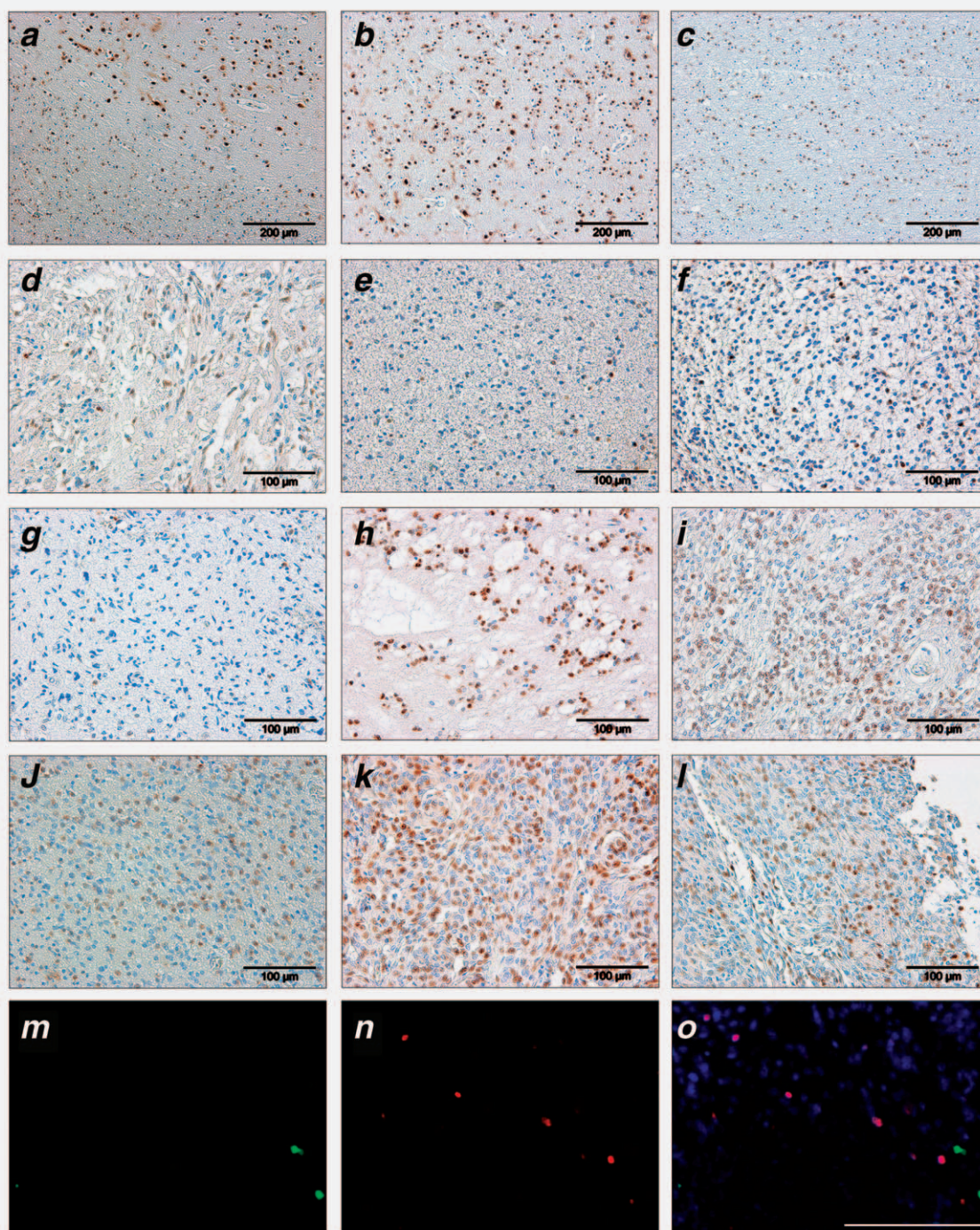


Figure 1. IHC with an antibody against 5hmC. (a) In the normal frontal lobe we see differences in staining of the cortex and subcortical white matter. In the cortex (upper right corner) there are numerous clearly visible positively stained neurons. In the subcortical white matter (lower left corner) there is diminished staining for 5hmC, but distinct positively stained cells are seen surrounded by negative ones. (b) In the normal occipital cortex, there is a clear staining of the majority (about 60%) of cells. (c) In normal occipital white matter, there are fewer 5hmC positive cells (about 30%) compared to the occipital cortex. (d) In human pilocytic astrocytomas (WHO Grade I) the majority of tumor cells are negative, only about 10% of tumor cells stain positive. (e) In diffuse astrocytomas less than 5% of cells stain positive for 5hmC. (f) AAs show lower values of 5hmC positive cells compared to low-grade gliomas. (g) GBMs show only very low positive staining for 5hmC. (h) In Grade I ependymomas most cells stain negative for 5hmC. (i) Ependymomas Grade II show lower numbers of 5hmC positive cells than Grade I tumors with a large intratumoral variation. (j) Anaplastic ependymomas show almost the same number of 5hmC positive cells compared to Grade II ependymomas but the cells stain much weaker. Furthermore there is a large intratumoral variation. (k) In Grade I meningiomas we see only a minority of positively stained cells with a large intratumoral variation. (l) Meningiomas Grade II show fewer 5hmC positive cells than Grade I tumors and the cells seem to stain more weakly than in Grade I tumors with large intratumoral variations. (m–o) Immunofluorescence staining of a GBM with antibodies against 5hmC (m), the proliferation marker Ki-67 (n) and DAPI. Scale bar is 100 μm. Cells stained positive for Ki-67 show no detectable signal for 5hmC in the merge picture (o).

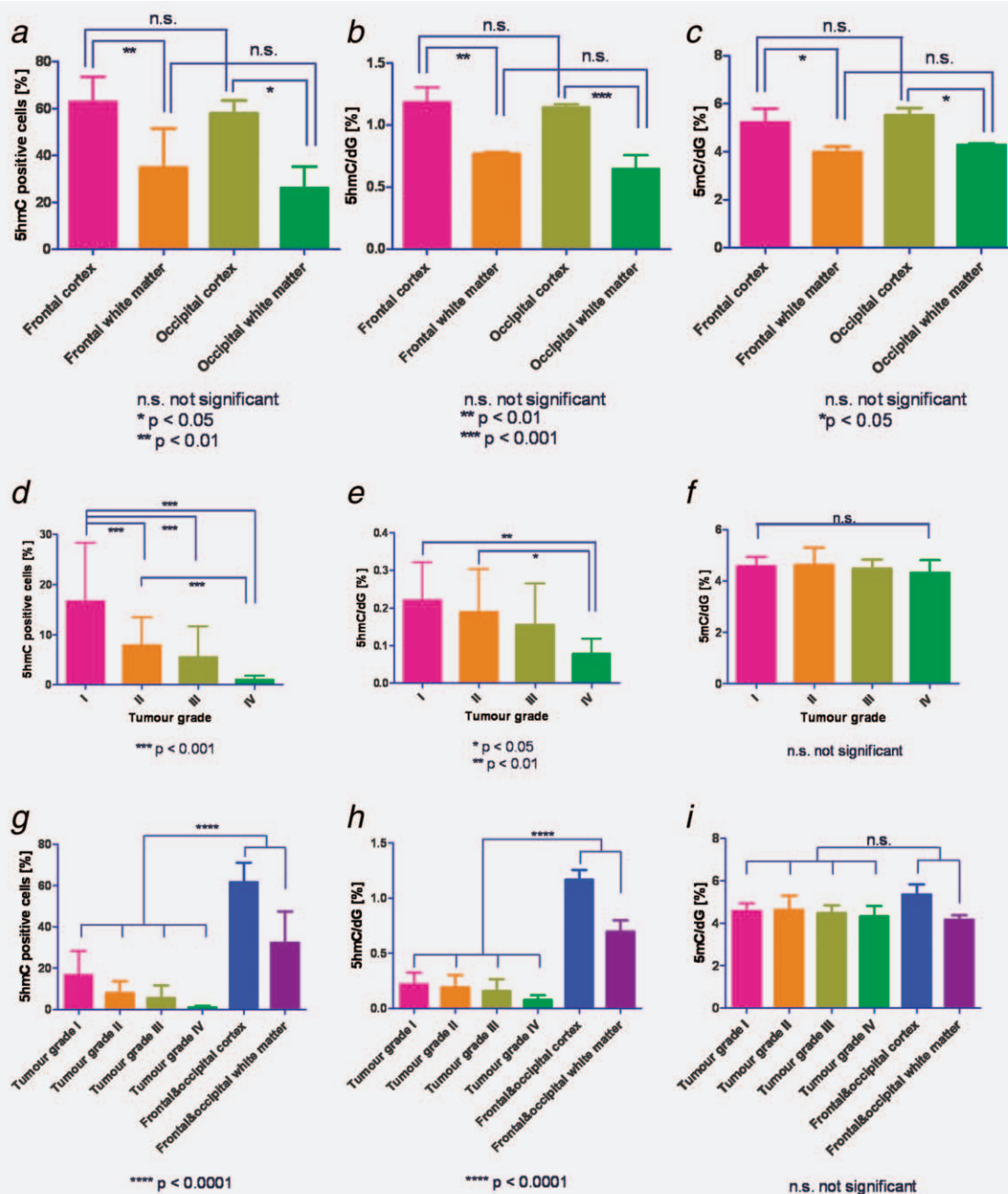


Figure 2. Percentage of 5hmC positive cells and LC-MS values of 5hmC and 5mC in normal human brain and in brain tumors. Indicated are the mean and SD. (a) There was no significant differences of 5hmC positive cells in various white matter areas as well as in various cortex regions of the human brain ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test) but there were highly significant differences in the cortex compared to the white matter ($p < 0.01$ in case of frontal cortex and frontal white matter and $p < 0.05$ in case of occipital cortex and occipital white matter, ANOVA and Tukey–Kramer *post hoc* test). (b) LC-MS values showed no significant difference between various white matter regions and between various cortex regions ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test) but highly significant differences between cortex and white matter regions ($p < 0.01$ in case of frontal cortex and frontal white matter and $p < 0.001$ in case of occipital cortex and occipital white matter, ANOVA and Tukey–Kramer *post hoc* test). (c) The distribution of 5mC in different brain regions showed no significant difference between different cortical regions as well as in different white matter regions ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test) but there were significant difference between cortical and white matter regions ($p < 0.05$, ANOVA and Tukey–Kramer *post hoc* test). (d) A comparison of 5hmC positive cells in human brain tumors showed highly significant differences between Grade I and Grade II tumors ($p < 0.001$, ANOVA and Tukey–Kramer *post hoc* test), Grade I and Grade III tumors ($p < 0.001$, ANOVA and Tukey–Kramer *post hoc* test), Grade I and Grade IV tumors ($p < 0.001$, ANOVA and Tukey–Kramer *post hoc* test) as well as Grade II and Grade IV tumors ($p < 0.001$, ANOVA and Tukey–Kramer *post hoc* test). (e) In LC-MS we found significant lower amounts of 5hmC in Grade I compared to Grade IV tumors ($p < 0.01$, ANOVA and Tukey–Kramer *post hoc* test). (f) The 5mC values did not show a correlation with tumor grade ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test). (g) A comparison of tumors with normal brain showed highly significantly lower numbers of 5hmC positive cells in tumors compared to normal brain ($p < 0.0001$, unpaired *t*-test). (h) In LC-MS we found that tumors show significant higher amount of 5hmC compared to normal brain ($p < 0.0001$, unpaired *t*-test). (i) There was no significant difference between tumor and normal brain in regard to 5mC ($p > 0.05$, unpaired *t*-test). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

By mass spectrometry we see that in Grade I tumors the 5hmC values were 0.22% 5hmC/dG, in Grade II 0.19% 5hmC/dG, in Grade III 0.16% 5hmC/dG and in Grade IV 0.078% 5hmC/dG with a significant difference between Grades I and IV tumors ($p < 0.01$, ANOVA and Tukey–Kramer *post hoc* test) as well as Grades II and IV tumors ($p < 0.05$, ANOVA and Tukey–Kramer *post hoc* test) (Fig. 2e).

No such relation was found in LC-MS measurements of 5mC in WHO tumor grades (Fig. 2f). Grade I tumors showed mean values of 4.58% 5mC/dG, Grade II 4.62% 5mC/dG, Grade III 4.47% 5mC/dG and Grade IV 4.32% 5mC/dG with no significant difference between different tumor grades ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test).

The number of 5hmC positive cells identified in brain tumors was significantly lower compared to normal cortex and white matter (Fig. 2g). While in normal human brain mean number of 5hmC positive cells was 61.48% in the cortex and 32.39% in the white matter, the mean in tumor tissue was 6.58% ($p < 0.0001$, unpaired *t*-test).

These data were paralleled by LC-MS (Fig. 2h). In normal cortex, we measured 1.17% 5hmC/dG, in the white matter 0.70% 5hmC/dG and in tumor tissue the average value was 0.16% 5hmC/dG ($p < 0.0001$, unpaired *t*-test).

As regards 5mC no significant difference was observed between normal brain tissue and tumor (Fig. 2i) with a mean of 4.85% 5mC/dG in the normal human brain and 4.50% 5mC/dG in tumors ($p > 0.05$, unpaired *t*-test).

When tumors were grouped according to their cellular lineages, we obtained similar differences of 5hmC related to the tumor grade.

In astrocytomas, the mean of 5hmC positive stained cells was 8.91% in Grade I pilocytic astrocytomas, 4.40% in Grade II diffuse astrocytomas, 3.34% in Grade III AAs and 1.07% in Grade IV GBMs (Fig. 3a) with highly significant differences between Grade I and Grade III tumors ($p < 0.001$, ANOVA and Tukey–Kramer *post hoc* test), Grade I and Grade IV tumors ($p < 0.001$, ANOVA and Tukey–Kramer *post hoc* test) and Grade II and Grade IV tumors ($p < 0.001$, ANOVA and Tukey–Kramer *post hoc* test) as well as significant differences between Grade I and Grade II tumors ($p < 0.01$, ANOVA and Tukey–Kramer *post hoc* test) and Grade III and Grade IV tumors ($p < 0.01$, ANOVA and Tukey–Kramer *post hoc* test) (Fig. 3a).

Using LC-MS, we found mean values of 0.19% 5hmC/dG in Grade I astrocytomas, 0.24% 5hmC/dG in Grade II, 0.13% 5hmC/dG in Grade III and 0.078% 5hmC/dG in Grade IV tumors (Fig. 3b) with significant differences between Grade II and Grade IV tumors ($p < 0.01$, ANOVA and Tukey–Kramer *post hoc* test) (Fig. 3b).

In regard to 5mC, there were no significant differences related to WHO grades of astrocytic gliomas using LC-MS (Fig. 3c). The mean value of Grade I astrocytomas was 4.68% 5mC/dG, of Grade II 4.82% 5mC/dG, of Grade III 4.36% 5mC/dG and of Grade IV 4.32% 5mC/dG ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test) (Fig. 3c).

In ependymal tumors, we found a tendency of 5hmC positive cells between Grade I as well as Grade II and Grade III tumors by IHC staining but we did not find significant differences ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test) (Fig. 3d). Grade I ependymomas showed mean values of 21.62% positive stained cells, Grade II and Grade III tumors show almost equal values of 5hmC positively stained cells (12.47% and 13.12%).

LC-MS measurements showed that WHO Grade I ependymomas have mean values of 0.286% 5hmC/dG, Grade II ependymomas 0.197% 5hmC/dG and anaplastic ependymomas (WHO Grade III) 0.193% 5hmC/dG and thus paralleled the same tendency as the immunohistochemical stainings ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test) (Fig. 3e).

With regard to 5mC no significant differences were found using LC-MS; in Grade I ependymomas the mean was 4.70% 5mC/dG, in Grade II 4.73% 5mC/dG and in Grade III 4.65% 5mC/dG ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test) (Fig. 3f).

A comparison of WHO Grades I and II meningiomas in IHC showed lower numbers of 5hmC positive cells in Grade II meningiomas compared to Grade I meningiomas (mean of 18.05% positive cells in Grade I and 6.51% positive cells in Grade II) but the difference did not reach statistical significance ($p > 0.05$, unpaired *t*-test) (Fig. 3g).

By mass spectrometry, we found a mean value of 0.163% 5hmC/dG in Grade I and 0.140% 5hmC/dG in one Grade II atypical meningioma, which again paralleled the same tendency as in IHC but this did not reach statistical significance ($p > 0.05$, unpaired *t*-test) (Fig. 3h).

Again, 5mC/dG levels did not show significant differences between Grade I and Grade II meningiomas (4.20% 5mC/dG in Grade I meningiomas and 4.32% 5mC/dG in Grade II meningiomas, $p > 0.05$, unpaired *t*-test) (Fig. 3i).

The number of 5hmC positive cells is significantly related to IDH1 mutations in diffuse and anaplastic astrocytomas but not in glioblastomas

Additionally, we analyzed the IDH1 status of all astroglial tumors by IHC using an IDH1^{R132H} antibody and pyrosequencing of the IDH1 gene. As expected we did not find IDH1 mutations in pilocytic astrocytomas (0% mutated IDH1), but we found 7 IDH1 mutations in diffuse astrocytomas (54% mutated IDH1), 9 IDH1 mutations in AAs (50% mutated IDH1) and 11 IDH1 mutations in GBMs (26% mutated IDH1).

Comparing 5hmC positive cells of diffuse and AAs as well as in GBMs (Fig. 4a) we found highly significant differences between diffuse astrocytomas with and without IDH1 mutation (1.50% 5hmC positive cells in tumors with IDH1 mutation compared to 7.31% 5hmC positive cells in tumors without IDH1 mutation, $p < 0.001$, ANOVA and Tukey–Kramer *post hoc* test) and significant differences in AAs with and without IDH1 mutation (1.80% 5hmC positive cells in tumors with IDH1 mutation and 4.87% 5hmC positive cells in tumors without IDH1 mutation, $p < 0.01$, ANOVA and Tukey–Kramer *post hoc* test). In GBM, we did not find significant differences

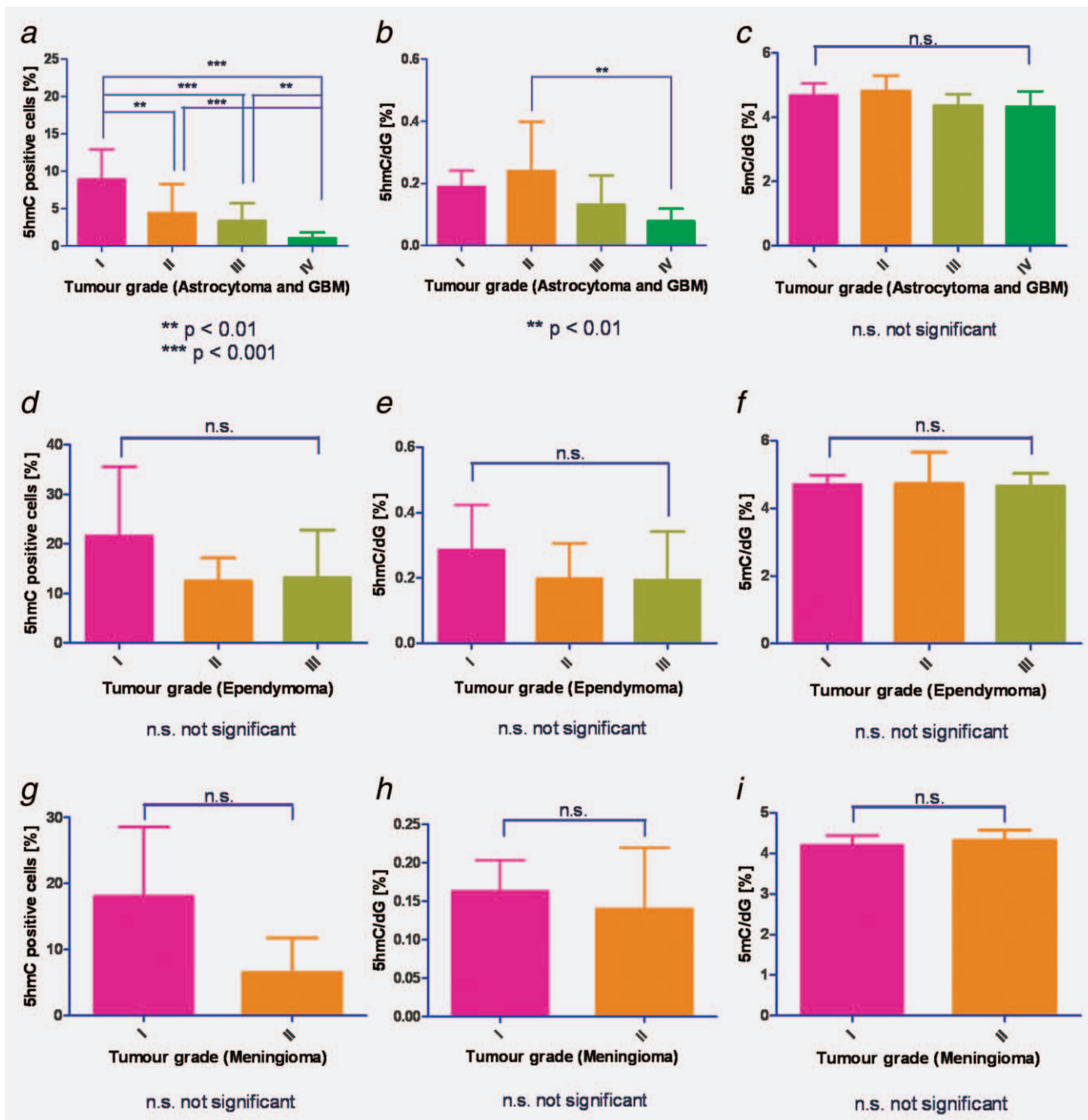


Figure 3. Percentage of 5hmC positive cells and LC-MS values of 5hmC and 5mC in various tumor entities. Indicated are the mean and SD. (a) In astrocytomas and GBMs a good correlation between the number of 5hmC positive cells and tumor grade was seen. Pilocytic Grade I astrocytomas showed highly significantly higher numbers of 5hmC positive cells compared to Grade III tumors ($p < 0.001$, ANOVA and Tukey–Kramer *post hoc* test) and Grade IV tumors ($p < 0.001$, ANOVA and Tukey–Kramer *post hoc* test), and Grade II tumors showed highly significant higher numbers of 5hmC positive cells compared to Grade IV GBM ($p < 0.001$, ANOVA and Tukey–Kramer *post hoc* test). Furthermore Grade I astrocytomas showed significantly lower numbers of 5hmC positive cells compared to Grade II tumors ($p < 0.01$, ANOVA and Tukey–Kramer *post hoc* test) as well as Grade III tumors showed significantly lower numbers of 5hmC positive cells compared to Grade IV GBM ($p < 0.01$, ANOVA and Tukey–Kramer *post hoc* test). (b) LC-MS showed similar results. Diffuse Grade II astrocytomas showed significantly higher amounts of 5hmC compared to undifferentiated Grade IV GBMs ($p < 0.01$, ANOVA and Tukey–Kramer *post hoc* test). (c) 5mC showed no significant differences between different tumor grades ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test). (d) Using IHC, in ependymomas Grade I there were more 5hmC positive cells than in ependymomas Grades II and III, but there was no significant decrease ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test). (e) Using LC-MS we found higher amounts of 5hmC in Grade I ependymomas compared to Grades II and III tumors without statistical significance ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test). (f) 5mC showed no significant differences in ependymomas of different WHO grades ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test). (g) Meningiomas showed a decrease of 5hmC positive cells in Grade II tumors compared to Grade I tumors but the difference did not reach statistical significance ($p > 0.05$, unpaired *t*-test). (h) LC-MS showed higher amounts of 5hmC in Grade I tumors compared to Grade II tumors without statistical significance ($p > 0.05$, unpaired *t*-test). (i) The amount of 5mC did not show any significant differences in regard to tumor grade ($p > 0.05$, unpaired *t*-test) [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

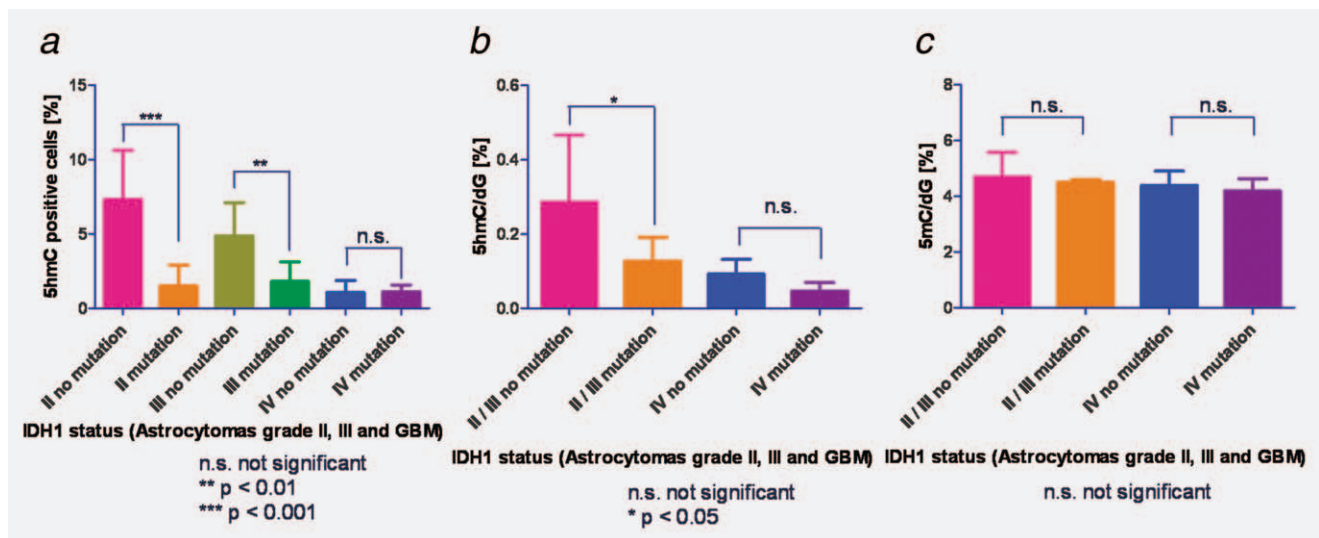


Figure 4. Number of 5hmC positive cells and LC-MS values of 5hmC/dG and 5mC/dG in astrocytomas and GBMs with regard to IDH1 status. Indicated are the mean and SD. (a) Comparing the number of 5hmC positive cells in diffuse and AAs as well as in GBMs with and without IDH1 mutations we found that diffuse astrocytomas without IDH1 mutation had significantly more 5hmC positive cells ($p < 0.001$, ANOVA and Tukey–Kramer *post hoc* test). Similar and highly significant differences were noted between AAs with and without IDH1 mutation ($p < 0.01$, ANOVA and Tukey–Kramer *post hoc* test). GBMs with and without IDH1 mutation did not show significant differences in the number of 5hmC positive cells ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test). Remarkably, Grade II diffuse and Grade III AAs that had a IDH1 mutation and GBMs with and without mutation had very low 5hmC counts, which were not significantly different between these groups ($p > 0.05$). (b) Using MS-LC we could parallel these results with significant differences between Grades II and III astrocytomas with and without mutant IDH1 ($p < 0.05$, ANOVA and Tukey–Kramer *post hoc* test). There was no significant difference in GBMs with and without IDH1 mutation ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test). (c) The amount of 5mC did not show any significant differences in regard to IDH1 mutations ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

between tumors with or without IDH1 mutation ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test) (Fig. 4a).

By mass spectrometry, we found a mean value of 0.287% 5hmC/dG in Grade II and Grade III tumors without IDH1 mutation and 0.127% 5hmC/dG in tumors with IDH1 mutation with a significant difference ($p < 0.05$, ANOVA and Tukey–Kramer *post hoc* test) (Fig. 4b). In GBM, we did not find significant differences between tumors with or without IDH1 mutation ($p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test) (Fig. 4b).

We did not find significant differences in the amount of 5mC/dG between Grades II and III astrocytomas with and without IDH1 mutation as well as GBMs (4.69% 5mC/dG in Grades II and III tumors without IDH1 mutation and 4.50% 5mC/dG in tumors with IDH1 mutation, $p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test, and 4.38% 5mC/dG in GBMs without IDH1 mutation and 4.19% 5mC/dG with IDH1 mutation, $p > 0.05$, ANOVA and Tukey–Kramer *post hoc* test) (Fig. 4c).

Discussion

The profoundly distorted epigenetic landscape of tumor cells is only beginning to emerge.^{12–14} DNA methylation, a major epigenetic mechanism of gene silencing, seems to play a major role in tumor pathogenesis. Transcriptional inactivation caused by promoter hypermethylation affects genes

involved in the main cellular pathways, DNA repair, vitamin response, Ras signaling, cell-cycle control, p53 network and apoptosis.¹⁵ Hypermethylation patterns are tumor-type specific and it is still unclear why certain regions become hypermethylated, whereas others remain unmethylated. Hypomethylation at specific promoters can activate the aberrant expression of oncogenes and induce loss of imprinting.¹⁴

DNA methylation is mediated by the DNMT family of enzymes that catalyze the transfer of a methyl group from S-adenosyl methionine to DNA. However, it seems that the DNA methylome is reprogrammed during development and we hypothesized that this might also be the case in gliomas. The discovery that 5mC can be oxidized to 5hmC by the action of the TET enzymes has engendered ideas of a role of TET and 5hmC in genetic reprogramming, *i.e.*, that 5hmC could be an intermediate leading to demethylation of 5mC and consequently a reversal of gene silencing by methylation. Indeed, very recent research has shown that TET proteins can convert 5mC to 5-formylcytosine and 5-carboxylcytosine and that the latter can be excised by thymine-DNA glycosylase (TDG), which has been interpreted as a pathway for active DNA demethylation.^{5–7}

The results presented here show that hydroxymethylation of genomic DNA cytidine residues is a feature found in the human central nervous system (CNS). The highest values were found in cortical areas (mean 1.17% 5hmC/dG) while

the white matter showed much lower values of 0.696% 5hmC/dG (ratio 1.68:1). Since neurons in the cortex are outnumbered by glia cells by a factor of 2–10 and IHC did not show a difference in the percentage of glial nuclei in the cortex and WM, we have to assume that the values of 5hmC/dG in neurons must be even higher. Percentage values of 5hmC/dG in the human adult cortex are the highest values published to date. The values of 5mC were around 5.3% of dG in the cerebral cortex and 4.2% in the cerebral white matter (ratio 1.26:1). Thus, overall it seems that a higher percentage of DNA is hydroxymethylated in neurons in absolute terms but also in relation to 5mC.

By comparison in the adult mouse, 5hmC/dG levels were highest in the CNS¹⁶; the adult murine cerebral cortex and hippocampus showed values around 0.6% 5hmC/dG while in other organs such as the kidney, heart and lung values were around 0.15%–0.2% and even lower in the liver, spleen and testis. Remarkably, 5mC/dG percentages were around 4.5% in most organs¹⁶ and various areas of the brain.¹⁷ Using an immunoprecipitation technique Jin *et al.* reported that 5hmC in DNA from human frontal lobe was more selectively targeted to genes than was 5mC and was particularly enriched at promoters and in intragenic regions. The presence of 5hmC was more positively correlated with gene expression than was the presence of 5mC.¹⁸

Surprisingly, our immunohistochemical results demonstrate that 5hmC is quite unevenly distributed in cells of neuronal and glial lineage. We would not conclude from our data that DNA in some cells is free of 5hmC, but our results unequivocally show that there is a wide variation in cellular 5hmC concentration.

In our investigation, 5hmC values in gliomas range from 0.086% to 0.24% 5hmC/dG and thus are surprisingly low when compared to measurements in the white matter. 5mC/dG values in tumors were around 4.5% and thus were slightly higher than in normal white matter. Since tumor infiltration areas could be a possible confounding factor that is difficult to exclude in investigating gliomas, we took great care to sample pure tumor areas excluding infiltration zones of the white matter or cortex as far as that is possible on light-microscopic inspection. However, tumor infiltration may still not be a totally negligible factor. When all tumor entities including astrocytomas, ependymomas and meningiomas were grouped by WHO grade, significantly lower 5hmC/dG values were noted in high-grade anaplastic tumors than in low-grade tumors while no significant differences were found in 5mC/dG values. When considering 5hmC/dG values of the individual tumor lineages, significant differences were found for low-grade astrocytomas *versus* GBMs, whereas only a tendency could be shown regarding ependymomas and meningiomas. Again, as in the normal brain, IHC demonstrated uneven cellular distribution in various tumor cell populations. We found that staining for 5hmC and the proliferation-associated Ki-67 was mutually exclusive and often observed a rhythmic variation of areas with higher proliferative activity (Ki-67) and sparse 5hmC staining and *vice*

versa. Again, 5mC/dG measurements showed no significant differences between WHO low- and high-grade tumors.

Thus by using LC-MS we have shown that 5hmC values related to the number of dG in the human brain are highest in gray matter areas, considerably lower in the white matter and even lower in brain tumors. Comparable values of 5hmC in the prefrontal cortex and a similar reduction in astrocytomas have recently been published by Jin *et al.*¹⁹; data on the white matter, GBMs and other brain tumors are not available at present. In the tumor group, lower 5hmC values were associated with anaplasia (WHO Grade IV), *i.e.*, a low degree of cellular differentiation and high mitotic activity. IHC with an antibody against 5hmC, more precisely the ratio of positively stained nuclei related to unstained nuclei, showed the same tendency as LC-MS; higher values were measured in the gray matter *vs.* white matter and highly differentiated tumors *vs.* anaplastic tumors. Additional factors such as the different cells of origin of these tumors may influence the 5hmC values. Pilocytic astrocytomas may be a particular case in point since they are known to be genetically different from higher-grade astrocytomas.

Acute myeloid leukemia (AML)-associated mutations of IDH1 and IDH2 that display a neomorphic enzyme activity resulting in the production of 2-hydroxyglutarate (2HG) have been shown to impair the catalytic function of TET2, an α -ketoglutarate (2-oxoglutarate or 2OG)-dependent enzyme.²⁰ Xu *et al.* reported decreased 5hmC values in a mixture of Grade III astrocytomas and GBMs harboring mutant IDH1.²¹ Similarly our data show in greater depth a significant difference in the percentage of 5hmC positive cells in diffuse and AAs with and without IDH1 mutation but not in GBMs (Fig. 4a). LC-MS measurements showed the same relationship between diffuse and AAs with and without IDH1 mutation and in GBMs.

Jin *et al.* found 5hmC levels were independent of IDH1 mutations in their sample of gliomas¹⁹; their data showed the same tendency as ours but did not reach statistical significance.

In a cohort of patients with AML, IDH1/2 mutations and TET2 loss-of-function mutations were mutually exclusive and elicited similar epigenetic defects.²⁰ Kim *et al.* found no TET2 mutation in 29 low-grade diffuse gliomas,²² neither have we identified TET2 loss-of function mutations in seven investigated cases of astrocytomas and GBMs (data not shown). Thus, concluding from what we know from myeloid leukemia, in human gliomas the observed very low 5hmC values may be related to functional impairment of the TET enzymes by 2HG. Remarkably, Grade II diffuse and Grade III AAs and GBMs that had IDH1 mutations had almost identical low 5hmC values. GBMs with no IDH1 mutation, which by and large are primary GBMs, had equally low values; therefore other, unknown factors may impair 5mC oxidation in these tumors.

Acknowledgements

The authors thank the Brain Bank Munich (Sigrun Roeber) for providing control tissues as well as Virginie Guibourt and Michael Schmidt for expert technical assistance. M. Mün. and T.P. thank the Fonds der Chemischen Industrie for pre-doctoral fellowships.

References

- Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, Kros JM, Hainfellner JA, Mason W, Mariani L, Bromberg JE, Hau P, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med* 2005;352:997–1003.
- Wick W, Hartmann C, Engel C, Stoffels M, Felsberg J, Stockhammer F, Sabel MC, Koeppen S, Ketter R, Meyermann R, Rapp M, Meisner C, et al. NOA-04 randomized phase III trial of sequential radiochemotherapy of anaplastic glioma with procarbazine, lomustine, and vincristine or temozolomide. *J Clin Oncol* 2009;27:5874–80.
- Kaina B, Christmann M, Naumann S, Roos WP. MGMT: key node in the battle against genotoxicity, carcinogenicity and apoptosis induced by alkylating agents. *DNA Repair (Amst)* 2007;6:1079–99.
- Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 2009;324:929–30.
- He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, Ding J, Jia Y, Chen Z, Li L, Sun Y, Li X, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 2011;333:1303–7.
- Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, He C, Zhang Y. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 2011;333:1300–3.
- Pfaffeneder T, Hackner B, Truss M, Munzel M, Muller M, Deiml CA, Hagemeier C, Carell T. The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew Chem Int Ed Engl* 2011;50:7008–12.
- Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 2009;324:930–5.
- Koh KP, Yabuuchi A, Rao S, Huang Y, Cuniff K, Nardone J, Laiho A, Tahiliani M, Sommer CA, Mostoslavsky G, Lahesmaa R, Orkin SH, et al. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* 2011;8:200–13.
- Haffner MC, Chaux A, Meeker AK, Esopi DM, Gerber J, Pellakuru LG, Toubaji A, Argani P, Iacobuzio-Donahue C, Nelson WG, Netto GJ, De Marzo AM, et al. Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. *Oncotarget* 2011;2:627–37.
- Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A, Scheithauer BW, Kleihues P. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathol* 2007;114:97–109.
- Esteller M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* 2007;8:286–98.
- Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. *Carcinogenesis* 2009;31:27–36.
- Portela A, Esteller M. Epigenetic modifications and human disease. *Nat Biotechnol* 2010;28:1057–68.
- Esteller M. Epigenetic gene silencing in cancer: the DNA hypermethylome. *Hum Mol Genet* 2007;16 Spec No 1:R50–9.
- Globisch D, Munzel M, Muller M, Michalakakis S, Wagner M, Koch S, Bruckl T, Biel M, Carell T. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* 2010;5:e15367.
- Munzel M, Globisch D, Bruckl T, Wagner M, Welzmler V, Michalakakis S, Muller M, Biel M, Carell T. Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew Chem Int Ed Engl* 2010;49:5375–7.
- Jin SG, Wu X, Li AX, Pfeifer GP. Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nucleic Acids Res* 2011;39:5015–24.
- Jin SG, Jiang Y, Qiu R, Rauch TA, Wang Y, Schackert G, Krex D, Lu Q, Pfeifer GP. 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Res* 2011;71:7360–5.
- Figueroa ME, Abdel-Wahab O, Lu C, Ward PS, Patel J, Shih A, Li Y, Bhagwat N, Vasanthakumari A, Fernandez HF, Tallman MS, Sun Z, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* 2010;18:553–67.
- Xu W, Yang H, Liu Y, Yang Y, Wang P, Kim SH, Ito S, Yang C, Xiao MT, Liu LX, Jiang WQ, Liu J, et al. Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell* 2011;19:17–30.
- Kim YH, Pierscianek D, Mittelbronn M, Vital A, Mariani L, Hasselblatt M, Ohgaki H. TET2 promoter methylation in low-grade diffuse gliomas lacking IDH1/2 mutations. *J Clin Pathol* 2011;64:850–2.

4.9 “Tet Oxidizes Thymine to 5-Hydroxymethyluracil in Mouse Embryonic Stem Cell DNA”

T. Pfaffeneder[#], F. Spada[#], **M. Wagner[#]**, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljaroova, D. Schürmann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schär, M. Müller, T. Carell, *Nat Chem Biol* **2014**, *10*, 574-581.

Co-first authors

Bei dem zurzeit weitgehend akzeptierten Mechanismus der aktiven DNA-Demethylierung treten die Basen fC und caC als Intermediate der aktiven DNA-Demethylierung auf. Sie entstehen bei der TET-katalysierten Oxidation von mC bzw. hmC und können im Rahmen der Basenexzisionreparatur von der Glykosylase TDG aus der DNA herausgeschnitten und durch unmodifizierte Cytosine ersetzt werden. Bevor dieser Mechanismus allgemein akzeptiert wurde, war eine Zeitlang vermutet worden, dass die Desaminierung von hmC zu 5-Hydroxymethyluracil (hmU) ein Schlüsselschritt der aktiven DNA-Demethylierung sein könnte. Deshalb sollte von uns zunächst ermittelt werden, wo und in welchen Mengen hmU in den Organen und Geweben der Maus und in murinen embryonalen Stammzellen vorkommt. Dies sollte die Grundlage für eine Überprüfung der Desaminierungshypothese sein, was u.a. durch *Isotope-tracing*-Experimente geschehen sollte.

In dieser Publikation haben wir eine verbesserte und sensitivere Methode zur Quantifizierung von modifizierten DNA-Nukleosiden vorgestellt. Mit dieser UHPLC-MS/MS-basierten Isotopenverdünnungsmethode konnten wir zeigen, dass die Desaminierung von hmC nur eine marginale Rolle bei der Bildung des genomischen hmUs spielt. Tatsächlich entsteht das *Steady-State*-Level von hmU in der DNA von embryonalen Stammzellen hauptsächlich durch die TET-katalysierte Oxidation von Thymin. In den Organen und Geweben adulter Individuen wird das genomische hmU dagegen größtenteils durch eine oxidative DNA-Schädigung gebildet.

Für diese Publikation habe ich in Zusammenarbeit mit Caterina Brandmayr die Experimente mit HEK-Zellen durchgeführt. Dabei habe ich Transfektionen und *Isotope-tracing*-Experimente vorgenommen, DNA-Proben isoliert, UHPLC-MS/MS-Messungen durchgeführt und die erhaltenen Daten ausgewertet. Diese Experimente lieferten den ersten Hinweis darauf, dass die TET-Enzyme neben den Cytosinderivaten auch Thymin oxidieren können.

Die UHPLC-MS/MS-basierte Quantifizierung der DNA-Modifikationen in den verschiedenen Organen und Geweben der Maus wurde von mir und Jessica Steinbacher durchgeführt. Daneben war ich an der Abfassung des Manuskriptes beteiligt und habe hierfür auch Grafiken angefertigt.

Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA

Toni Pfaffeneder^{1,8}, Fabio Spada^{1,8}, Mirko Wagner^{1,8}, Caterina Brandmayr¹, Silvia K Laube¹, David Eisen¹, Matthias Truss², Jessica Steinbacher¹, Benjamin Hackner¹, Olga Kotljarova¹, David Schuermann³, Stylianos Michalakos⁴, Olesea Kosmatchev¹, Stefan Schiesser¹, Barbara Steigenberger¹, Nada Raddaoui¹, Gengo Kashiwazaki¹, Udo Müller⁵, Cornelia G Spruijt⁶, Michiel Vermeulen^{6,7}, Heinrich Leonhardt⁵, Primo Schär³, Markus Müller^{1*} & Thomas Carell^{1*}

Ten eleven translocation (Tet) enzymes oxidize the epigenetically important DNA base 5-methylcytosine (mC) stepwise to 5-hydroxymethylcytosine (hmC), 5-formylcytosine and 5-carboxycytosine. It is currently unknown whether Tet-induced oxidation is limited to cytosine-derived nucleobases or whether other nucleobases are oxidized as well. We synthesized isotopologs of all major oxidized pyrimidine and purine bases and performed quantitative MS to show that Tet-induced oxidation is not limited to mC but that thymine is also a substrate that gives 5-hydroxymethyluracil (hmU) in mouse embryonic stem cells (mESCs). Using MS-based isotope tracing, we show that deamination of hmC does not contribute to the steady-state levels of hmU in mESCs. Protein pull-down experiments in combination with peptide tracing identifies hmU as a base that influences binding of chromatin remodeling proteins and transcription factors, suggesting that hmU has a specific function in stem cells besides triggering DNA repair.

Methylcytosine is an epigenetically important nucleobase associated with the control of transcriptional activity, genomic imprinting, X-chromosome inactivation and suppression of transposable elements¹. Controlled formation and removal of mC at specific genomic loci is critical for correct genome programming or reprogramming during cellular differentiation². Recently, it was discovered that Tet proteins (Tet1–3) oxidize mC to give the oxidized C-derived nucleobases hmC³, 5-formylcytosine (fC)^{4,5} and 5-carboxycytosine (caC)^{5,6}, whose biological functions are still yet unclear (Fig. 1a)⁷. As fC and caC are both removed by thymine DNA glycosylase (Tdg)^{6,8}, it is currently assumed that they serve as intermediates of an active DNA demethylation process involving base excision repair. In addition to these oxidized C derivatives, cells also contain oxidized T nucleobases such as hmU and fU. These compounds are currently known as oxidative lesions that are thought to form upon the reaction of T with reactive oxygen species (ROS)^{9,10}. It was recently suggested that hmU might also be produced by deamination of hmC, a hypothesis that remains controversial^{11–14}. Deamination of hmC, situated in a base pair with G (hmC:G), would give rise to hmU:G mismatches, which are known substrates for the DNA glycosylases Tdg, Smug1, Mbd4, Ung2 (ref. 15), Neil1 and Nthl1 (ref. 16). Deamination of hmC:G to hmU:G followed by mismatch repair would therefore establish an alternative pathway to active demethylation (Fig. 1a).

To unravel the origin of oxidized nucleobases, and of hmU in particular, in DNA from mESCs, we performed isotope tracing and quantitative MS studies using the chemically synthesized

isotopologs of mC, hmC, fC, caC, hmU and fU as internal standards (Fig. 1b and Supplementary Results, Supplementary Fig. 1). For the assessment of oxidation products that are formed by the action of ROS, we additionally quantified 8-oxo-G because 8-oxo-G is a well-established ROS reaction product formed from G^{17,18}. We show here that hmU is generated enzymatically from thymidine during stem cell differentiation by the action of the Tet enzymes. A proteomic analysis provides new insight into how genomic hmU can influence the binding of chromatin remodeling proteins and transcription factors.

RESULTS

hmU is present at elevated levels in mESCs

We first created an inventory of the named nucleosides (Supplementary Fig. 1) in mESCs (Fig. 1c) and adult cortex tissue (Fig. 1d). In mESCs, we observed that hmC is, as expected, the most abundant oxidized pyrimidine (10% relative to mC), followed by fC (~1–2% of hmC). The ROS marker 8-oxo-G was detected at similar levels (~45% of fC), showing that nonenzymatic, ROS-induced oxidations of nucleobases are important processes, as expected. We also found relatively high levels of fU (22% of fC). Clearly detectable were also caC and hmU. hmU, which is at the center of this study, was unequivocally detected, as shown by its retention time and its specific fragmentation pattern, which were found to be identical with that of the internal standard [D₂]hmU (Fig. 1e). Notably, both caC and hmU were present in comparable amounts (5% of fC). In adult mouse cortex DNA (Fig. 1d; for other tissues see

¹Center for Integrated Protein Science at the Department of Chemistry, Ludwig-Maximilians-Universität München, München, Germany. ²Charité Universitätsklinikum, Otto-Heubner-Centrum für Kinder und Jugendmedizin, Klinik für Allgemeine Pädiatrie, Labor für Pädiatrische Molekularbiologie, Berlin, Germany. ³Department of Biomedicine, University of Basel, Basel, Switzerland. ⁴Center for Integrated Protein Science at the Department of Pharmacy—Center for Drug Research, Ludwig-Maximilians-Universität München, München, Germany. ⁵Center for Integrated Protein Science at the Department of Biology, Ludwig-Maximilians-Universität München, Planegg-Martinsried, Germany. ⁶Department of Molecular Cancer Research, Cancer Genomics Netherlands, Utrecht, The Netherlands. ⁷Present address: Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, Nijmegen, The Netherlands. ⁸These authors contributed equally to this work.

*e-mail: markus.mueller@cup.uni-muenchen.de or thomas.carell@cup.uni-muenchen.de

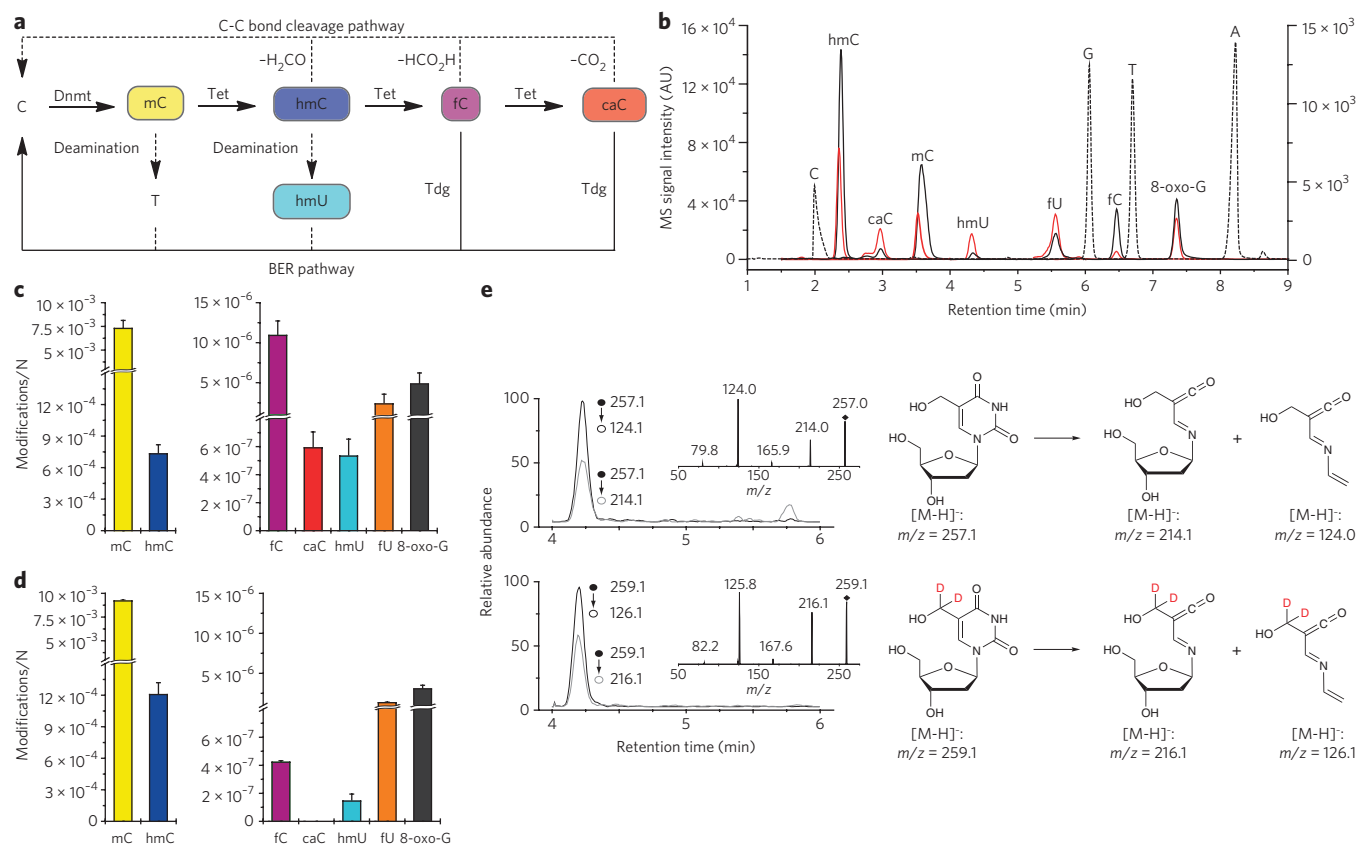


Figure 1 | Metabolism of cytosine derivatives, their detection by LC-UV-ESI-MS/MS and levels in mESCs and mouse cortex. (a) Potential active demethylation pathways. (b) Overlaid LC/UV and LC/MS/MS chromatograms of a representative DNA sample from mESCs. The dotted LC/UV chromatogram of C, G, T and A is scaled arbitrarily; the overlaid LC/MS/MS chromatograms of hmC, mC, fC and 8-oxo-G are scaled to the left y axis; the LC/MS/MS chromatograms of caC, hmU and fU are scaled to the right y axis. Red chromatograms refer to the corresponding labeled internal standards depicted in **Supplementary Figure 1**. AU, arbitrary units. (c,d) DNA modification levels per nucleoside (N) in mESCs (c; WT01, $n = 7$) and 3-month-old mouse cortex tissue (d; $n = 3$). Depicted are mean values \pm s.d. (e) Representative LC/MS/MS chromatograms for identification and quantification of hmU. Shown are the overlaid chromatograms for the two characteristic fragment ions of hmU (top trace) and the [D₂]hmU internal standard (bottom trace) derived from a mESC DNA sample. The insets show the MS/MS full-scan spectra of synthetic hmU and [D₂]hmU matching the proposed fragmentation pathway.

Supplementary Fig. 2), the hmC levels are very high (13% relative to mC), as previously reported^{19–21}. In contrast, fC was detected only in very small amounts (0.03% relative to hmC), and caC was not observed at all²². The detected amounts of 8-oxo-G and fU are comparable. Notably, the detected levels of hmU in the cortex were significantly lower compared to that in mESCs (27%; $P = 2.2 \times 10^{-4}$ by Student's t -test). Because the detected levels of 8-oxo-G were similar in the cortex (**Fig. 1d**) and in mESCs (**Fig. 1c**), the elevated hmU levels in mESCs cannot be explained by greater ROS-induced damage in mESCs and must have a different origin.

We next investigated this in more detail and quantified the levels of hmU in three different mESC lines (WT01, J1 and R1) and in a variety of tissues (**Fig. 2a**). Clearly, the hmU levels were higher in mESCs than in somatic tissues by factors of 2–15. The detected hmU levels correspond roughly to 500–1,700 hmU bases per genome in tissue and 2,900–7,800 in mESCs (**Fig. 2a**) depending on the cell type and growth conditions. We cannot explain these differences by elevated oxidative stress levels in mESCs (**Supplementary Fig. 3**). We next analyzed the levels of C, mC and the oxidized pyrimidines hmC, fC, hmU and fU as well as 8-oxo-G in somatic tissues (**Supplementary Fig. 2**) and performed a correlation and cluster analysis to reveal relationships of the modification levels (**Fig. 2b**). The data analysis confirmed that the low levels of hmU in somatic tissues correlate strongly (significant on a level <0.001) with the levels of fU and 8-oxo-G within a separated cluster.

This result showed that in somatic cells, hmU and fU are ROS-induced reaction products. If we assume that in somatic tissues hmU is exclusively formed by ROS, the data showed that in mESCs, 70–80% of the detected hmU is produced by ROS-independent processes (**Supplementary Fig. 3**).

hmU formation by oxidation of T

To analyze the origin of the oxidized nucleobases in mESCs, we performed isotope tracing experiments (**Fig. 3a–c** and **Supplementary Fig. 4**). Substitution of L-methionine with [methyl-¹³CD₃]L-methionine ([¹³CD₃]L-Met) in the growth medium is known to give the labeled S-adenosylmethionine cofactor, which is needed for the conversion of C to mC. Replacement of L-Met by [¹³CD₃]L-Met for 5 d (2 passages) in the medium at a concentration of 0.2 mM furnished 89% labeled [¹³CD₃]mC, 88% labeled [hydroxymethyl-¹³CD₂]hmC and 93% labeled [formyl-¹³CD]fC (**Fig. 3b** and **Supplementary Fig. 4**). Within the detection limit (7 and 50 molecules per 10⁸ nucleosides, respectively), no incorporation of the isotopes ¹³C and D into hmU and fU was observed, showing that hmC is not the precursor of hmU. We next added isotope-labeled [¹³C,¹⁵N₂]thymidine ([¹³C,¹⁵N₂]T) to the growth medium and observed ~76% of label incorporation into T, hmU and fU (**Fig. 3a**, **Supplementary Fig. 4** and **Supplementary Table 1**). The combined data showed that hmU is not generated by deamination of hmC but by oxidation of T (T → hmU; **Fig. 3c**). Consequently, all of the detected hmU

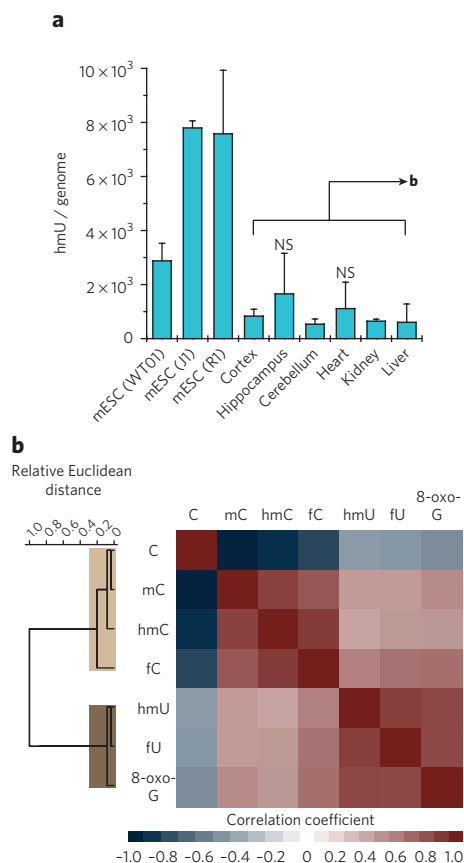


Figure 2 | hmU is present at elevated levels in mESCs compared to tissue.

(a) hmU levels per genome in mESCs ($n_{\text{WT01}} = 7$ replicates, $n_{\text{H}} = 2$, $n_{\text{R1}} = 3$) and mouse tissue (3-month-old individuals, $n = 3$). Levels per genome were obtained considering a mouse genome size of 2.7×10^9 base pairs. Depicted are mean values \pm s.d. The differences between mESCs and mouse tissues are significant ($P = 2.9 \times 10^{-5}$ to 3.6×10^{-2} ; unpaired two-tailed t -test) except for WT01 and hippocampus ($P = 0.292$) or heart ($P = 0.069$). These exceptions are due to higher hmU levels in the hippocampus and heart caused by higher background oxidation (higher 8-oxo-G levels). hmU levels normalized to oxidative background (8-oxo-G) levels are in **Supplementary Figure 3**. NS, not significant. (b) Unsupervised clustering analysis of Pearson correlation coefficients of 24 data sets of selected mouse organs at a defined time point (3-month-old individuals: cortex, hippocampus, cerebellum, heart, liver and kidney). Discussed correlations are strong to very strong (Pearson coefficient >0.7) and significant on a level <0.001 .

(and also fU) resides in an A base pair context (hmU:A). In this base pair, hmU is repaired by Smug1 but not Tdg¹². This was confirmed by siRNA-mediated knockdown of Smug1 in mESC cells and HEK-293T cells. Indeed, the hmU level increased in these cells (**Supplementary Fig. 5**).

The fact that we were unable to detect hmU derived from hmC deamination, which would be situated in a base pair with G (hmU:G), could also be explained by very fast repair. If repair of the hmU:G base pair is extremely efficient, for example, because deamination and glycosylase-based repair occurs in a tight complex of the involved enzymes, we would be unable to detect this type of hmU because of low steady state levels. Indeed, it was proposed that hmC deamination and hmU excision requires a complex of the cytidine deaminase Aid and the glycosylase Tdg¹². To assess this possibility, we performed isotope tracing experiments using [¹³CD₃]_L-Met in Tdg^{-/-} mESCs stably complemented with either empty vector (control) or a minigene expressing a catalytically incompetent Tdg at near-endogenous levels. These cells are able to form the Aid-Tdg

complex, but the Tdg is inactive, which should give elevated hmU levels if deamination occurs. In both cell lines, we detected high levels of labeled hmC. In the control cells, [hydroxymethyl-¹³CD₂] hmU was not detected. However, in the cell line complemented with inactive Tdg, some labeled [hydroxymethyl-¹³CD₂]hmU was indeed observed, albeit only in small amounts ($\sim 7\%$ of total hmU and $\sim 0.06\%$ of total hmC; **Supplementary Fig. 6**). Thus, the Tdg protein is required for deamination of hmC to hmU. In wild-type (WT) mESCs, the hmU:G mismatches are obviously repaired so quickly that they do not contribute to steady state levels of hmU. All of the detected hmU was derived from T oxidation and resided in hmU:A base pairs.

Tet enzymes form hmU in correlation with mC oxidation

We next investigated whether enzymatic oxidation or ROS-dependent processes are responsible for the observed oxidation of T to hmU. To this end, we studied mESCs with genetic or functional depletions of Tet or DNA methyltransferase (Dnmt) enzymes (**Fig. 4**). Although hmU levels were maintained at normal levels in the severely hypomethylated, Dnmt-depleted cell lines, they were substantially reduced upon knockdown of Tet1 and Tet2. As the levels of fU and 8-oxo-G remained unaffected, the data establish that, though both hmU and fU are generated by T oxidation, their formation must occur by two independent processes in mESCs. The data supported the idea that fU is a ROS-created lesion similar to 8-oxo-G^{9,10,17,18}, whereas most hmU is generated by Tet-induced oxidation of T.

In support of enzymatic T oxidation by Tet proteins, ectopic expression of the catalytic domain of Tet1 (Tet1cd) in HEK-293T cells led to a 65-fold increase for both hmC and hmU levels. This was not observed when a catalytic mutant of Tet1 (Tet1cm) was expressed (**Supplementary Fig. 7a**). To determine whether under these conditions hmU is generated by deamination, as previously suggested^{11,12}, we again replaced natural _L-Met with [¹³CD₃]_L-Met in the medium, but we did not detect incorporation of heavy isotopes into hmU, providing evidence that the elevated hmU levels do not originate from hmC deamination (**Supplementary Table 1**). Finally, we confirmed enzymatic generation of hmU *in vitro* by incubating recombinant Tet1cd with a plasmid that was premethylated by bacterial methyltransferase M.SssI. In addition to oxidation of mC to hmC, fC and caC, we detected hmU at a prominent level of 9% relative to hmC (**Supplementary Fig. 7b**), showing that the catalytic center of the Tet enzymes clearly has the capacity to oxidize T to hmU.

To further confirm that hmU is formed in mESCs in the process of epigenetic reprogramming, we analyzed the dynamic changes of mC and all of the oxidized pyrimidines plus 8-oxo-G during differentiation. It was recently shown that mC and hmC levels sharply increase when mESCs maintained in the naive state are shifted to a primed state in serum-containing medium^{23–25}. To investigate global kinetics of all of the Tet-generated oxidation products under more physiologically relevant priming conditions, we used established protocols based on serum-free N2B27 medium for differentiation of naive mESCs into states resembling that of post-implantation epiblasts^{26,27}. mESCs were first grown for several passages in the presence of MEK and GSK3 inhibitors (dual inhibition or 2i conditions) and LIF to induce a hypomethylated state resembling that of the naive epiblast^{23–25}. The data in **Figure 5a** are averaged from three independent differentiation experiments, each performed with two cell lines in the absence of growth factors. First, we observed that the fU and 8-oxo-G levels stayed constant (**Fig. 5a**), in line with the idea that both are ROS-derived products. In contrast, fC and caC levels peaked at about 8 h. Both hmU and hmC also peak between 8 h and 16 h. Isotope tracing with [¹³CD₃]_L-Met under these conditions provided no evidence for switched-on deamination of hmC to hmU (**Supplementary Table 1**). At these peak levels, we estimated that mESCs contain roughly 110,000 fC bases, 4,400 caC bases and

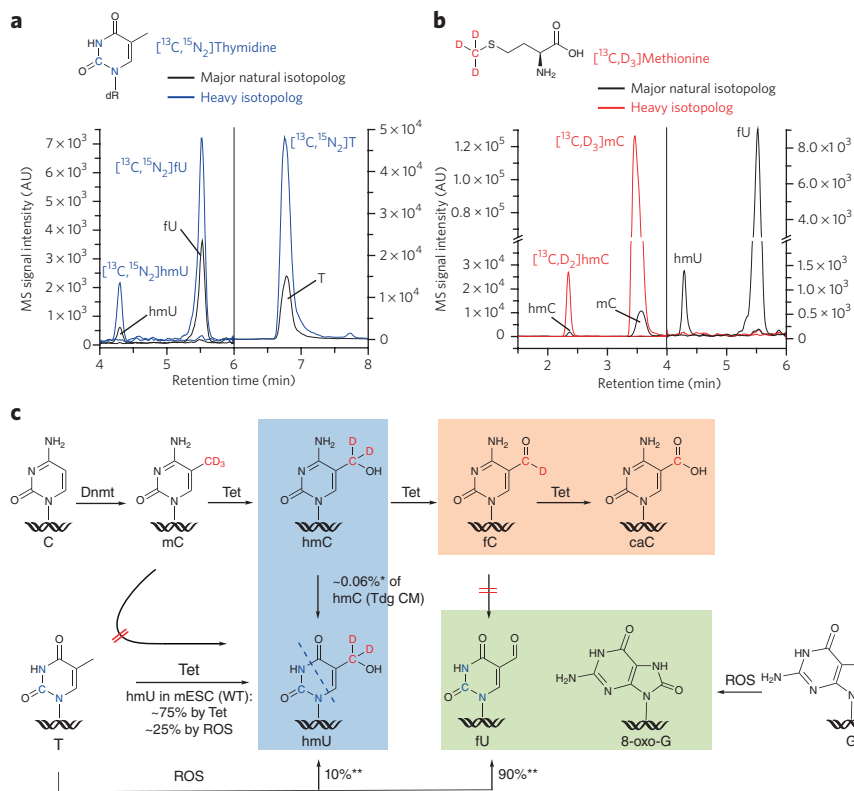


Figure 3 | hmU and fU are thymine oxidation products in WT mESCs with no detectable contribution from hmC or fC deamination. (a) Overlaid LC/MS/MS chromatograms of heavy ($^{13}\text{C}_5, ^{15}\text{N}_2$)hmU, $^{13}\text{C}_5, ^{15}\text{N}_2$ fU and $^{13}\text{C}_5, ^{15}\text{N}_2$ T; (blue) and major natural (black) isotopologs of hmU, fU and T. AU, arbitrary units. (b) Overlaid chromatograms of heavy ($^{13}\text{C}_5, \text{D}_3$)hmC, $^{13}\text{C}_5, \text{D}_3$ fC, $^{13}\text{C}_5, \text{D}_3$ hmU and $^{13}\text{C}_5, \text{D}_3$ fU; (red) and the major natural (black) isotopologs of mC, hmC, hmU and fU. (c) Enzymatic and ROS-dependent pathways leading to the formation of hmC, fC, caC, hmU and fU. Single asterisks denote labeled hmU generated by deamination of labeled hmC, which was observed only in Tdg catalytic mutant (CM) cells (Supplementary Fig. 6), representing ~7% of the total hmU content and corresponding to deamination of ~0.06% hmC. Double asterisks denote basal rates of ROS-dependent T oxidation, which were determined in HEK-293T cells, where Tet activity is lowest (3.6 hmC per 10^5 nucleosides; Supplementary Table 2). Here, hmU and fU roughly represent 10% and 90% of T oxidation products relative to the sum of each other.

14,000 hmU bases per genome, showing that, at its peak level, hmU is three times more abundant than caC. The dynamic peaking data allowed us to estimate half-life times for fC (7 h), caC (5 h) and hmU (4 h) during the differentiation process (Supplementary Fig. 8). Using the dynamic quantitative data, we performed a correlation and cluster analysis of the DNA modification levels (Fig. 5b). Three independent clusters were obtained. One cluster involves members of cytosine methylation (and demethylation) dynamics, including C, mC, fC and caC, but, to our surprise, not hmC. Instead, hmC groups with hmU. We therefore concluded that the formation of these modifications is tightly coupled by the action of the Tet enzymes. The ROS-induced lesions fU and 8-oxo-G form the third, well-separated cluster. Notably, hmU does not correlate with the ROS lesions.

Recently, it was shown that exposure of naive mESCs to fibroblast growth factor 2 (FGF-2) and activin A (ActA) for 48 h under similar conditions as those described above gives a homogeneous cell population whose transcriptome closely resembles that of the post-implantation epiblast²⁶. Under these conditions, mC levels increased more rapidly, approaching somatic levels within a time frame closely reflecting that observed during embryonic development (Supplementary Fig. 9a). Levels of hmC rose steadily throughout the 48-h time course. Whereas fC peaked at 36 h, hmU reached its maximum at 24 h (Supplementary Fig. 9b). qPCR data

showed a robust induction of Dnmt3b that is most likely responsible for the increasing mC levels (Supplementary Fig. 9a). A transient peak of Tet1 expression at 24 h went in hand with rising hmU and fC levels, whereas Tet3 expression started rising slowly from the same time point and may drive further accumulation of hmC and fC (Supplementary Fig. 9b). Tdg was progressively upregulated together with the onset of methylation activity, whereas Smug1 showed little, if any, change (Supplementary Fig. 10). These data supported our view that hmU peaking is not caused by downregulation of the repair pathway.

Thus, the kinetic data showed a complex interplay between methylation and two oxidation reactions during differentiation, which depends on the exact conditions. Most important is the fact that hmU showed a time-dependent occurrence similar to the other oxidized bases hmC, fC and caC independently of the conditions investigated, confirming its formation during epigenetic reprogramming.

hmU attracts specific readers

To obtain initial insight into a potential biological function of hmU in comparison to hmC, we screened for specific readers associating with hmU:A as well as with hmC:G and hmU:G using protein pull-down and relative quantification by LC/MS/MS. In previous stable isotope labeling by amino acids in cell culture (SILAC)-based proteomics studies with hmC, fC and caC containing oligonucleotides, we and others observed a high number of specific protein readers, arguing that the new bases influence a variety of different processes^{28,29}. For this hmU study, we further developed this approach for the detection of proteins that directly interact with the modified bases hmU (as well as hmC for comparison) to get a more direct insight into their function (Fig. 6a). For the study, we not

only included into the biotinylated DNA duplexes (24mers) an hmU (hmC) base but also equipped the counter strand with a polyethylene glycol-based linker carrying a reactive *N*-hydroxysuccinimide (NHS) ester moiety and a reductively cleavable disulfide bond in the middle. This reactive linker cannot be inserted into the DNA strand using solid phase synthesis. In addition, the NHS ester does not survive hybridization conditions. We therefore attached the linker as its azide derivative, using Cu(I)-catalyzed click chemistry, to an alkyne-bearing base present in the DNA duplex^{30,31}. These DNA duplexes were subsequently incubated with nuclear extracts from mESCs. The DNA-bound protein complexes were isolated using streptavidin-coated magnetic beads (Fig. 6a)³². DNA duplexes with the canonical base pairs A:T and C:G at the respective positions served as reference strands. The NHS linker has two functions in the experiment: First, it will covalently trap the reader proteins specifically at the ϵ -amino groups of lysines, which allows the identification of transiently binding proteins as well. Second, because the linker is cleavable, it will tag the trapped lysine residues, leaving a defined label on those proteins that bind in close proximity to the modified bases hmC and hmU. This allowed us to distinguish protein readers that bound close to hmC and hmU from proteins that are secondary members of the complexes. The covalently trapped proteins were next tryptically digested and labeled with tandem mass tagging (TMT) isobaric tags to allow protein

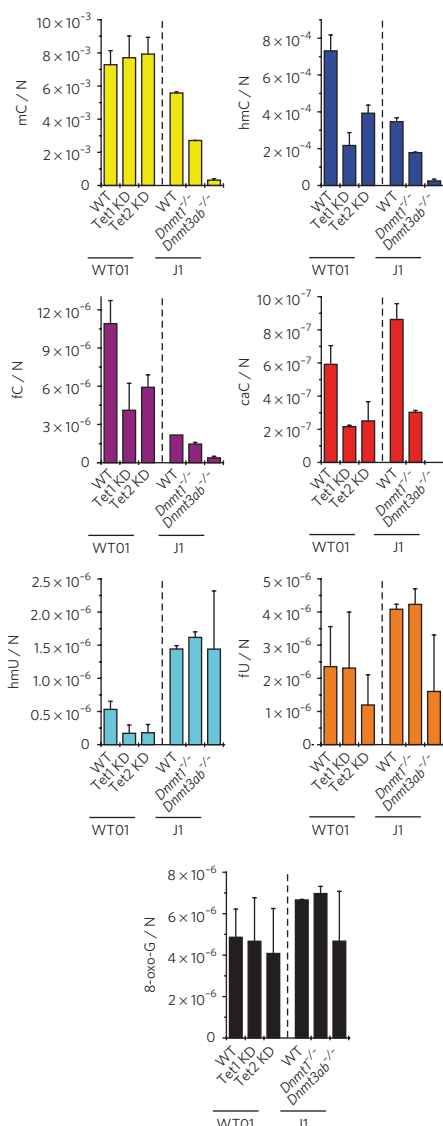


Figure 4 | Tet1 and Tet2 generate hmU in mESCs. Effect of Tet and Dnmt depletion on DNA modification levels in mESCs. J1 cell lines bearing homozygous *Dnmt*-null mutations and WT01 mESCs expressing Tet1 or Tet2 shRNAs (knockdown (KD)) were analyzed. Depicted are mean values per nucleoside (N) ± s.d. of biological replicates as follows: wild type (WT01, *n* = 7), Tet1 KD (WT01, *n* = 3) and Tet2 KD (WT01, *n* = 3); WT (J1, *n* = 2), *Dnmt1*^{-/-} (J1, *n* = 2) and *Dnmt3a/b*^{-/-} (J1, *n* = 6).

identification and quantification by MS^{32,33}. Only proteins that were enriched in both the forward and the reverse TMT experiment were considered to be specific hmU (hmC) readers (Fig. 6). All of the readers were subsequently divided into two groups. Proteins that were enriched relative to the control strands are termed specific readers (sRs). Of those, the proteins that were identified with a peptide containing the tag are termed direct-specific readers (dsRs).

The scatter plot (Fig. 6b and Supplementary Fig. 11a) revealed over 250 sRs recognizing the hmU:A base pair (relative to T:A). Of those proteins, 99 were identified with tagged peptides. These proteins were therefore designated dsRs. A gene ontology analysis showed that 64 proteins of the 250 sRs are nucleotide binders, and 25 proteins are involved in chromatin organization (Supplementary Fig. 11b).

Among the dsRs we identified were the regulatory proteins Uhrf2, transcription factor HIVP3 or the poly [ADP-ribose] polymerase 14

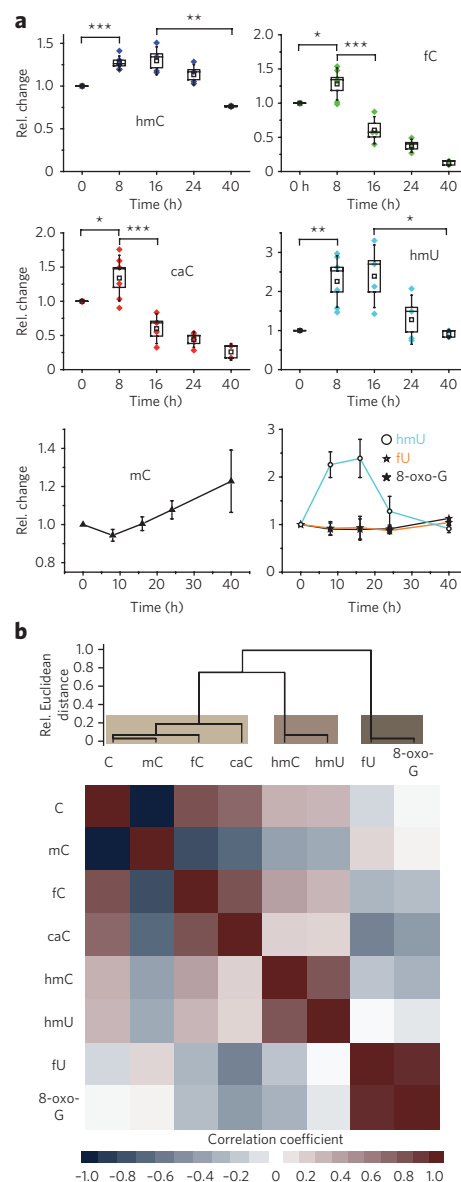


Figure 5 | hmU is produced during mESC differentiation. (a) Time course of DNA modification levels during early differentiation (0–40 h) of naive mESCs in the absence of growth factors. Box plot of relative modification changes for hmC, fC, caC and hmU averaged from six independent experiments using two different mESC lines. In the four upper plots, colored diamonds reflect mean values of technical triplicate measurements, open squares reflect mean values of biological replicates, boxes represent the s.e.m., and whiskers represent the s.d. of the biological replicates; **P* < 0.05, ***P* < 0.01 and ****P* < 0.005 by unpaired two-tailed *t*-test. In the two lower plots, the relative modification changes of mC, hmU, fU and 8-oxo-G are depicted as biological mean values ± s.e.m. (b) Unsupervised clustering analysis of Pearson correlation coefficients of 22 data sets obtained from the differentiation experiments depicted in a. Discussed correlations are strong to very strong (Pearson coefficient >0.7) and significant on a level <0.001.

(Parp14), which showed that these proteins interacted directly with hmU. We also detected several H2A and H2B histone proteins among the dsRs and the sRs. Notably, in the group of the sRs, we observed also Dnmt3a and Dnmt3b, which are involved in gene expression regulation. Methyl CpG-binding protein 2 (MeCP2) was identified as well, but it showed only low enrichment. We also detected several transcriptional regulators, such as bromodomain PHD finger

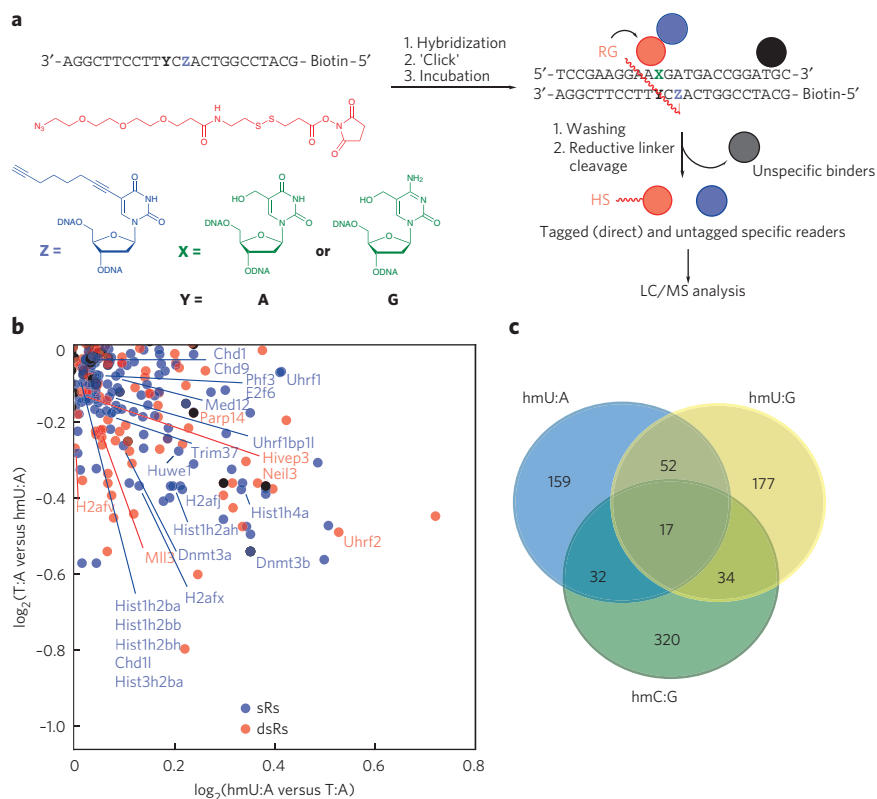


Figure 6 | Identification of hmU:A readers. (a) Workflow of pulldown experiments with hmU:A-containing oligonucleotides. DNA oligomers with the modified base Z containing an alkyne group for click reaction and biotin for the pulldown are hybridized with DNA oligomers containing hmU (X = hmU and Y = A). After the click reaction with the DNA-protein cross-linker (red), which has an NHS ester as a lysine-specific reactive group (RG), the dsDNA oligomers are incubated with a nuclear extract from mESCs, and specific binders are covalently linked to the DNA. After three washing steps, the linker is reductively cleaved, and the proteins are enzymatically digested and labeled with TMT2plex reagents. The same dsDNA oligomer without hmU modification (X = T and Y = A) was used as a control sample. Specifically enriched proteins are mixed with the control sample before LC/MS analysis. The same workflow was used for experiments with hmU:G or hmC:G and C:G as a control. (b) Magnified view of proteins enriched with hmU:A-containing oligomer. Specific readers in the forward and reverse experiment are marked in blue. Direct-specific readers are identified by the presence of the DNA-protein cross-linker and marked in red. Black dots are considered unspecific binders. A full-scale representation is shown in **Supplementary Figure 11**. (c) Venn diagram showing overlap of specific binders for hmU:A, hmU:G and hmC:G.

transcription factor (Bptf), transcription factor E2F6 and mediator of RNA polymerase II transcription subunit 12 (Med12). Moreover, we identified hmU-binding E3 ubiquitin-protein ligases such as Uhrf1, Trim37 and Huwe1 and finally the chromodomain helicase DNA-binding proteins 1 and 9 (Chd1 and Chd9) among the hmU readers, which are chromatin remodeling factors and regulate polymerase I and II transcription. To exclude bias by sequence-specific binding, we repeated the experiment with hmU in a different sequence context. Indeed, 78% of the identified protein readers are also observed in this control experiment, and included among them were all of the proteins discussed above, with the notable exception of Uhrf1, which was not observed in this experiment (**Supplementary Data Set 1**). The data show that hmU:A recruits specific proteins that are involved in chromatin biochemistry in a broad sense.

As a first validation of the MS-identified readers, we focused on Uhrf1 and Uhrf2. Uhrf1 is a known mC and hmC binder³⁴, whereas we recently identified the homolog Uhrf2 as a specific binder of hmC²⁹. The MS data for hmU obtained in this study showed an enrichment of Uhrf1 and Uhrf2. In addition, Uhrf2 was identified as a direct binder. To validate the MS findings, we overexpressed

Uhrf1 or Uhrf2 together with Tet1cd in HEK-293T²⁹ and measured the levels of mC, hmC, fC, caC and also hmU. Indeed, we observed increased levels of hmU, proving a functional relation between hmU levels and Uhrf1 and Uhrf2 (**Supplementary Fig. 12**). As Uhrf2 is known to interact with Dnmt3a and Dnmt3b³⁵ and these two proteins are also identified as sRs of hmU, our initial data provide a consistent picture.

Analysis of the specific hmU readers showed that 49 of the identified proteins were also enriched with an hmC-modified strand. Among these proteins were, for instance, bromodomain PHD finger transcription factor (Bptf), MeCP2 and 19 other nucleotide-binding proteins. The obtained list of specifically binding proteins is in good agreement with our recently published data set²⁹, which further supports the validity of the covalent trapping method. Finally, we compared the results of the hmU:A pulldown with a pulldown of hmU:G, which would be the product of hmC deamination. We found 69 of 280 sRs in both experiments. We identified transcription regulators as sRs as well in the hmU:G pulldown. However, we also found several DNA repair proteins, such as DNA repair protein complementing XP-C cells homolog (XPC), DNA repair protein Rad50 and five helicases, among others (**Supplementary Data Set 1**).

DISCUSSION

This work aimed to analyze the origin of hmU in mESCs. Although hmU as well as fU are both well-characterized oxidation products of thymidine^{9,10}, hmU in particular was recently discussed as an intermediate in active demethylation by deamination of hmC^{11,12}. We used stable isotopologs of the main oxidation products hmC, fC, caC, hmU, fU and 8-oxo-G in combination with quantitative MS to show that the levels of hmU are strongly elevated in stem cells in comparison to somatic tissue. Notably, we did not observe equally elevated levels of fU and 8-oxo-G, both of which are formed by

the reaction of DNA with ROS. This analysis suggests that hmU in stem cells is not only formed by ROS. The quantitative data allowed us to estimate that in WT stem cells, about 70–80% of the found hmU is not formed by ROS but is derived from a different process. Isotope tracing studies with labeled [¹³CD₃]_L-Met and [¹³C,¹⁵N₂] thymidine showed that the detected hmU originates exclusively from T. Previous studies already proposed a connection between Tet activity and hmU levels, but these studies did not address the origin of hmU, which led to the hypothesis that hmU could form by deamination of hmC^{11,22}. To investigate the involvement of deamination^{11–14}, we studied *Tdg*^{-/-} stem cells reconstituted with a catalytically inactive Tdg. In these cells, the putative deamination–repair complex¹² of Tdg and Aid could form, and we expected to see elevated hmU levels. Indeed, in this experiment, we were able to detect labeled hmU, showing that deamination of hmC does occur, but the levels were low. Most notably, special conditions were required to detect this deamination. In WT cells, all of the detected hmU is clearly derived from T oxidation. As such, the detected hmU is situated in a base pair with A (hmU:A). In conclusion, the majority of hmU in mESCs is produced independently of both ROS and deamination.

Using a combination of functional knockdown in mESCs, ectopic expression of Tet1 catalytic domain in HEK-293T cells and *in vitro* studies with recombinant Tet protein, we showed that the Tet enzymes, known to oxidize mC to hmC, are also responsible for oxidation from T to hmU. A kinetic study in which the temporal changes of all of the oxidized pyrimidines were observed under two different differentiation kinetics showed finally that hmU is formed in processes that also generate hmC and fC. Together with the observed peak levels of hmU of about 14,000 bases per genome, this suggests that hmU could have an epigenetic function similar to hmC. It is conceivable that the Tet enzymes introduce a small amount of hmU during the hmC manufacturing process to trigger faster demethylation by induction of DNA repair. It is known that hmC is not a substrate for repair^{8,12}, whereas hmU is efficiently recognized and repaired³⁶, for example, by the repair glycosylase Smug1 (refs. 37,38). Sporadic introduction of hmU could therefore allow recruitment of repair factors, for example, for long patch repair, as recently suggested³⁹. Alternatively, hmU might influence transcription factor binding, which is known to be a hallmark of epigenetic reprogramming⁴⁰. To study the proteins that closely interact with hmU, a new pulldown experiment was devised in which we inserted a trapping linker into the oligonucleotide. This linker is able to react covalently with proteins that assemble on the DNA duplex close to the hmU base. Because the linker is reductively cleavable, it leaves a tag on the respective lysine residues, which is detected in subsequent MS-based proteomics measurements. Using this technology, we observed that hmU:A recruits transcription factors and proteins that are involved in chromatin biochemistry, including Uhrf1 and Uhrf2. This result is in line with the recent observation that oxidative lesions in DNA can change the binding affinity of the transcription factor CREB when they are present in the respective cognate sequence⁴¹. A further analysis in HEK-293T revealed that the hmU readers Uhrf1 and Uhrf2 are able to modulate the levels of all of the oxidized mC bases and of hmU. It is interesting to note that the observed level increases of hmC were small, whereas they were substantial for hmU and also for the further oxidized bases fC and caC. As Uhrf1 and Uhrf2 are proteins known to be involved in numerous epigenetic processes, our data supported the hypothesis that Tet-induced oxidation of T to hmU may have an epigenetic function. Even in the case that hmU is formed just as a side product of Tet-induced hmC formation, it is now clear that the oxidation chemistry performed by the Tet enzymes has to go in hand with effective DNA repair.

Received 15 December 2013; accepted 17 April 2014;
published online 18 May 2014

METHODS

Methods and any associated references are available in the [online version of the paper](#).

References

- Smith, Z.D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).
- Franchini, D.M., Schmitz, K.M. & Petersen-Mahrt, S.K. 5-Methylcytosine DNA demethylation: more than losing a methyl group. *Annu. Rev. Genet.* **46**, 419–441 (2012).
- Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
- Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **50**, 7008–7012 (2011).
- Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
- He, Y.F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
- Tan, L. & Shi, Y.G. Tet family proteins and 5-hydroxymethylcytosine in development and disease. *Development* **139**, 1895–1902 (2012).
- Maiti, A. & Drohat, A.C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286**, 35334–35338 (2011).
- Bjelland, S. *et al.* Oxidation of thymine to 5-formyluracil in DNA: mechanisms of formation, structural implications, and base excision by human cell free extracts. *Biochemistry* **34**, 14758–14764 (1995).
- Mouret, J.F., Polverelli, M., Sarrazini, F. & Cadet, J. Ionic and radical oxidations of DNA by hydrogen peroxide. *Chem. Biol. Interact.* **77**, 187–201 (1991).
- Guo, J.U., Su, Y., Zhong, C., Ming, G.L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423–434 (2011).
- Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79 (2011).
- Nabel, C.S. *et al.* AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* **8**, 751–758 (2012).
- Rangam, G., Schmitz, K.M., Cobb, A.J. & Petersen-Mahrt, S.K. AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. *PLoS ONE* **7**, e43279 (2012).
- Jacobs, A.L. & Schär, P. DNA glycosylases: in DNA repair and beyond. *Chromosoma* **121**, 1–20 (2012).
- Zhang, Q.M. *et al.* DNA glycosylase activities for thymine residues oxidized in the methyl group are functions of the hNEIL1 and hNTH1 enzymes in human cells. *DNA Repair (Amst.)* **4**, 71–79 (2005).
- Burrows, C.J. Surviving an oxygen atmosphere: DNA damage and repair. *ACS Symp. Ser. Am. Chem. Soc.* **2009**, 147–156 (2009).
- Taghizadeh, K. *et al.* Quantification of DNA damage products resulting from deamination, oxidation and reaction with products of lipid peroxidation by liquid chromatography isotope dilution tandem mass spectrometry. *Nat. Protoc.* **3**, 1287–1298 (2008).
- Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5**, e15367 (2010).
- Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
- Münzel, M. *et al.* Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.* **49**, 5375–5377 (2010).
- Liu, S. *et al.* Quantitative assessment of Tet-induced oxidation products of 5-methylcytosine in cellular and tissue DNA. *Nucleic Acids Res.* **41**, 6421–6429 (2013).
- Ficz, G. *et al.* FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **13**, 351–359 (2013).
- Habibi, E. *et al.* Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360–369 (2013).
- Leitch, H.G. *et al.* Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* **20**, 311–316 (2013).
- Hayashi, K., Ohta, H., Kurimoto, K., Aramaki, S. & Saitou, M. Reconstitution of the mouse germ cell specification pathway in culture by pluripotent stem cells. *Cell* **146**, 519–532 (2011).
- Ying, Q.L., Stavridis, M., Griffiths, D., Li, M. & Smith, A. Conversion of embryonic stem cells into neuroectodermal precursors in adherent monoculture. *Nat. Biotechnol.* **21**, 183–186 (2003).
- Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14**, R119 (2013).
- Spruijt, C.G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
- Burley, G.A. *et al.* Directed DNA metallization. *J. Am. Chem. Soc.* **128**, 1398–1399 (2006).
- Rostovtsev, V.V., Green, L.G., Fokin, V.V. & Sharpless, K.B. A stepwise Huisgen cycloaddition process: copper(I)-catalyzed regioselective “ligation” of azides and terminal alkynes. *Angew. Chem. Int. Ed.* **41**, 2596–2599 (2002).
- Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895–1904 (2003); erratum **75**, 4942 (2003); erratum **78**, 4235 (2006).
- Gygi, S.P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **17**, 994–999 (1999).
- Frauer, C. *et al.* Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS ONE* **6**, e21306 (2011).
- Pichler, G. *et al.* Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J. Cell. Biochem.* **112**, 2585–2593 (2011).
- Lewis, H.L., Muhleman, D.R. & Ward, J.F. Serologic assay of DNA base damage. I. 5-Hydroxymethyldeoxyuridine, a radiation product of thymidine. *Radiat. Res.* **75**, 305–316 (1978).
- Boorstein, R.J. *et al.* Definitive identification of mammalian 5-hydroxymethyluracil DNA N-glycosylase activity as SMUG1. *J. Biol. Chem.* **276**, 41991–41997 (2001).

38. Kavli, B., Otterlei, M., Slupphaug, G. & Krokan, H.E. Uracil in DNA—general mutagen, but normal intermediate in acquired immunity. *DNA Repair (Amst.)* **6**, 505–516 (2007).
39. Santos, F. *et al.* Active demethylation in mouse zygotes involves cytosine deamination and base excision repair. *Epigenetics Chromatin* **6**, 39 (2013).
40. Silva, J. & Smith, A. Capturing pluripotency. *Cell* **132**, 532–536 (2008).
41. Moore, S.P.G., Toomire, K.J. & Strauss, P.R. DNA modifications repaired by base excision repair are epigenetic. *DNA Repair (Amst.)* **12**, 1152–1158 (2013).

Acknowledgments

We thank the Excellence Cluster Center for Integrated Protein Science Munich (CiPS^M) and the collaborative research centers SFB749, SFB646 and SFB1032 as well as German Research Foundation (DFG) grant CA275/8-4, the Volkswagen foundation, NGFNplus (01GS0870) and the Netherlands Organization for Scientific Research (NWO-VIDI) for financial support. T.P. and S.S. thank the Fonds der Chemischen Industrie for predoctoral fellowships. C.B. thanks the Boehringer Ingelheim Fonds for a predoctoral fellowship. G.K. thanks the Japan Society for the Promotion of Science (JSPS) for a postdoctoral fellowship for research abroad. We thank M. Moser (Max Planck Institute for Biochemistry) for providing R1- and C57Bl6/129-derived mESCs, G. Höfner and K.T. Wanner for their initial help with MS as well as M. Wirsing, L. Belzner and P. Laube for providing bioinformatic tools for data processing.

Author contributions

T.P. synthesized MS standards, performed the sample preparation and ultra high-performance LC/MS/MS method development, did LC/MS analysis, interpreted data and performed statistical analysis. F.S. and N.R. performed mESC differentiation and isotope tracing experiments. M.W. and C.B. performed the HEK-293T experiments, did LC/MS analysis and interpreted data. C.B. performed qPCR and analyzed the data. S.K.L. and D.E. performed the protein pulldown studies and interpreted data. M.T. performed mESC knockdown and knockout experiments. J.S. and O. Kosmatchev did sample preparation and LC/MS analysis. B.H., S.S. and J.S. prepared MS standards. O. Kotljarova performed *in vitro* assays. B.S. synthesized oligonucleotides for protein capture. G.K. synthesized tandem mass tags, and S.M. provided mouse tissue samples. U.M. and H.L. constructed Tet expression plasmids. C.G.S. and M.V. performed Uhrf1/2 overexpression in HEK-293T cells. P.S. and D.S. provided plasmids and cell lines. M.M. and T.C. conceived and supervised the project, interpreted data and wrote the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available in the [online version of the paper](#). Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Correspondence and requests for materials should be addressed to M.M. and T.C.

ONLINE METHODS

General materials and methods. Chemicals were purchased from Sigma-Aldrich, Fluka, ABCR or Acros Organics and used without further purification. Acetonitrile of LC/MS grade was purchased from Carl Roth GmbH & Co., KG. Formic acid, p.a. for MS, was purchased from Fluka, and water was purified with a Milli-Q Plus system from Merck Millipore.

The MS standards 5-methyl-2'-deoxycytidine (mC), 5-trideuteromethyl-2'-deoxycytidine ($[D_3]$ mC), 5-hydroxymethyl-2'-deoxycytidine (hmC), 5-dideuterohydroxymethyl-2'-deoxy- $(N^1,N^3-^{15}N_2)$ -cytidine ($[^{15}N_2,D_2]$ hmC), 5-formyl-2'-deoxycytidine (fC), 5-formyl-2'-deoxy- $(N^1,N^3-^{15}N_2)$ -cytidine ($[^{15}N_2]$ fC), 5-carboxy-2'-deoxycytidine (caC), 5-carboxy-2'-deoxy- $(N^1,N^3-^{15}N_2)$ -cytidine ($[^{15}N_2]$ caC), 5-hydroxymethyl-2'-deoxyuridine (hmU), 5-(dideuterohydroxymethyl)-2'-deoxyuridine ($[D_2]$ hmU), 5-formyl-2'-deoxyuridine (fU) and 5-formyl-2'-deoxy- $(N^1,N^3-^{15}N_2)$ -uridine ($[^{15}N_2]$ fU) were synthesized according to earlier published work^{19,21,42}. All of the synthesized compounds were characterized and purity confirmed by 1H -NMR, ^{13}C -NMR and ESI-MS, and some were additionally validated by ^{15}N -NMR. 8-hydroxy-2'-deoxy- $(^{15}N_5)$ -guanosine ($[^{15}N_5]$ 8-oxo-G) (99 atom% ^{15}N) was purchased from Cambridge Isotope Laboratories; 8-hydroxy-2'-deoxyguanosine (8-oxo-G) was from BIOLOG; 2'-deoxyguanosine (G) and 2'-deoxycytidine (C) were from ChemGenes. (methyl- $^{13}C,D_3$)-L-Met (99 atom% D and ^{13}C) was purchased from Sigma-Aldrich, and 2'-deoxy- $(C^2-^{13}C, N^1,N^3-^{15}N_2)$ -thymidine (99 atom% ^{13}C and ^{15}N) from Hartmann Analytic. Aqueous stock solutions of these compounds were stored at $-20^\circ C$ and warmed up to RT before usage.

Oligonucleotide synthesis. Oligonucleotide synthesis was performed on an ABI 394 DNA/RNA synthesizer (Applied Biosystems) using standard DNA synthesis conditions (DMT off) and acetyl-protected dC. Phosphoramidites (including the 5-hydroxymethyl-dU-CE phosphoramidite and the 5'-biotin phosphoramidite) and polystyrene carriers were obtained from Glen Research. 5-octadynyl-dU (Z) phosphoramidite was synthesized according to literature and characterized by 1H -NMR, ^{13}C -NMR, ^{31}P -NMR and ESI-MS⁴³. The crude oligonucleotide was cleaved from the resin and deprotected in 30% (v/v) ammonium hydroxide and 40% (v/v) methylamine (1:1) at $65^\circ C$ for 10 min and purified by preparative and analytical HPLC (Waters Breeze and Alliance, respectively). Separation was performed by applying a VP 250/10 Nucleosil 100-7 C18 column (flow: 5 ml/min) from Macherey-Nagel with a gradient of buffer A (0.1 M NH_4Et_3OAc in water) and buffer B (0.1 M NH_4Et_3OAc in 80% MeCN). DNA-containing fractions were characterized by MALDI-TOF (Bruker Autoflex II) and analytical HPLC, combined, and desalted by C18-Sep-Pak cartridges (Waters). For analytical HPLC, separation was performed by applying a CC 250/4 Nucleosil 120-3 C18 column from Macherey-Nagel (flow: 0.5 ml/min) with the aforementioned buffer system.

For the protein pulldown studies, an hmU-containing oligonucleotide was hybridized with a DNA strand modified with a 5'-biotin and a 5-octadynyl-dU. The sequences of the DNA strands are summarized in **Supplementary Table 3**. To 10 nmol of the dsDNA, 0.5 μ l of a 200 mM solution of the cross-linking azide-PEG₃-S-S-NHS ester ($C_{18}H_{29}N_5O_8S_2$, Jena Bioscience, Jena, Germany) in DMSO was added. In a separate tube, CuBr was dissolved in a TBTA solution (DMSO/*t*BuOH 3:1, 100 mM) resulting in a 1:1 Cu(I):TBTA ratio. This solution was immediately added to the DNA/azide mixture. Furthermore, 45 μ l of DMSO/*t*BuOH (3:1) were added, the mixture was shaken at $37^\circ C$ for 3 h, and the resulting product was purified by ethanol precipitation.

LC/MS analysis of DNA samples. Quantitative LC/UV-ESI-MS/MS analysis of digested DNA samples was performed using an Agilent 1290 UHPLC system equipped with a UV detector and an Agilent 6490 triple quadrupole mass spectrometer coupled with the stable isotope dilution technique. An improved method, based on earlier published work^{29,42,44,45}, was developed, which allowed the concurrent analysis of all nucleosides in one single analytical run. The source-dependent parameters were as follows: gas temperature $50^\circ C$, gas flow 15 l/min (N_2), nebulizer 30 psi, sheath gas heater $275^\circ C$, sheath gas flow 11 l/min (N_2), capillary voltage 2,500 V in the positive ion mode, capillary voltage $-2,250$ V in the negative ion mode and nozzle voltage 500 V. The fragmentor voltage was 380 V. Delta EMV was set to 500 (positive mode) and 800 (negative mode). Compound-dependent parameters are summarized in **Supplementary Tables 4** and **5**. Chromatography was performed by a Poroshell 120 SB-C8 column (Agilent, 2.7 μ m, 2.1 mm \times 150 mm) at $30^\circ C$ using a gradient of water

and MeCN, each containing 0.0085% (v/v) formic acid, at a flow rate of 0.35 ml/min: 0 \rightarrow 5 min; 0 \rightarrow 3.5% (v/v) MeCN; 5 \rightarrow 6.9 min; 3.5 \rightarrow 5% MeCN; 6.9 \rightarrow 7.2 min; 5 \rightarrow 80% MeCN; 7.2 \rightarrow 10.5 min; 80% MeCN; 10.5 \rightarrow 11.3 min; 80 \rightarrow 0% MeCN; 11.3 \rightarrow 13 min; 0% MeCN. The effluent up to 1.5 min and after 9 min was diverted to waste by a Valco valve. The autosampler was cooled to $10^\circ C$. The injection volume was amounted to 29 μ l. Calibration curves, method validation and data processing are in **Supplementary Note 2**. A complete compilation of LC/MS quantifications results see **Supplementary Note 3**.

DNA digestion. 5–25 μ g of genomic DNA in 25 μ l H_2O were digested as follows: An aqueous solution (7.5 μ l) of 480 μ M $ZnSO_4$, containing 42 U nuclease S1 (*Aspergillus oryzae*, Sigma-Aldrich), 5 U Antarctic phosphatase (New England BioLabs) and specific amounts of labeled internal standards (**Supplementary Note 2**) were added, and the mixture was incubated at $37^\circ C$ for 3 h. After addition of 7.5 μ l of a 520 μ M $[Na]_2$ -EDTA solution, containing 0.2 U snake venom phosphodiesterase I (*Crotalus adamanteus*, USB corporation), the sample was incubated for another 3 h at $37^\circ C$ and then stored at $-20^\circ C$. Prior to LC/MS/MS analysis, samples with up to 15 μ g DNA, for which the quantification of low amounts of caC was aspired, were filtered by using an AcroPrep Advance 96 filter plate 0.2 μ m Supor (Pall Life Sciences). In contrast, samples with 15–25 μ g DNA (isotope-tracing experiments) were filtered by using an AcroPrep Advance 96 filter plate 10K Omega (Pall Life Sciences).

Genomic DNA isolation. Tissues of female WT mice (C57-BL/6N) were dissected at postnatal day 90 and prepared as earlier described^{19,21}. Genomic DNA was extracted using the Qiagen Blood and Cell Culture DNA Midi Kit except for mESC samples differentiated in the presence of growth factors (see below). Extraction was performed following the manufacturer's instructions for genomic DNA isolation from cell culture samples or tissue samples, respectively. All buffers until loading of the sample on Genomic-tip 100/G were additionally supplemented with antioxidants 3,5-di-*tert*-butyl-4-hydroxytoluene (BHT, 200 μ M) and deferoxamine mesylate salt (desferal, 200 μ M) as well as the deaminase inhibitor tetrahydrouridine (THU, 200 μ M), according to published methods, to reduce background oxidation or deamination¹⁸. Elution buffer QF was supplemented with 200 μ M BHT. Following elution, all steps were performed on ice. DNA was then precipitated with NaOAc (0.3 M final) and 0.7 volumes *i*PrOH. DNA pellets from cultured cells were washed twice with ice-cold 70% EtOH and resuspended in H_2O containing 20 μ M BHT using a Qiagen TissueLyser (30 Hz, 2 min). DNA pellets from mouse tissues were resuspended in PBS buffer and additionally extracted with phenol/ $CHCl_3$, precipitated, washed and resuspended as described above.

R1 mESC samples differentiated in the presence of growth factors or transfected with Smug1 esiRNAs were lysed directly in the plates with RLT buffer (Qiagen) supplemented with BHT and desferal as described above. DNA was isolated using the Zymo Quick gDNA Midi Kit according to the manufacturer's instruction, except that elution was repeated four times with 100 μ l of elution buffer supplemented with BHT (200 μ M). Eluted DNA was precipitated with 2 M ammonium acetate and two volumes of absolute ethanol and finally resuspended in H_2O containing 20 μ M BHT. The flow-through from the spin columns was used to isolate RNA (see real-time PCR analysis).

mESC cell culture. Feeder independent WT01 mESCs (C57BL/6 strain)⁴⁶ were cultured in the presence of serum and LIF as previously described⁴. *Tdg*^{+/−} and *Tdg*^{−/−} mESCs were described previously⁴⁷. *Tdg*^{−/−} mESCs were complemented by random integration of either empty vector (hereafter referred to as *Tdg*^{−/−} mESCs) or a minigene expressing catalytically inactive Tdg (N151A)⁴⁷. Clonal mESC lines with targeted Tdg alleles, R1 cells (strain 129/Sv)⁴⁸, J1 cell lines (strain 129S4/SvJae)⁴⁹ and a mESC line derived from C57BL/6/129 mixed background⁵⁰ were routinely maintained on gelatinized plates in DMEM (PAA or Sigma) supplemented with 10% FBS, 1 \times MEM-nonessential amino acids (NEAA), 0.2 mM L-alanyl-L-glutamine, 100 U/ml penicillin, 100 μ g/ml streptomycin (all from PAA), 0.1 mM β -mercaptoethanol, 20 ng/ml ($\geq 1 \times 10^3$ U/ml) mouse recombinant LIF (ORF Genetics), 1 μ M PD 0325901 and 3 μ M CHIR 99021 (2i; both from Axon Medchem). In these conditions, the global levels of genomic mC were very low (and, as a consequence, the levels of its oxidized derivatives were even lower; data not shown). Before DNA isolation, 2i cultures were passaged twice (over 5 d) in DMEM supplemented with FBS and LIF as above but lacking 2i. With this strategy, primed mESC cultures were obtained with no sign of overt differentiation and modified genomic cytosines reached reproducibly higher and stable levels. For isotope tracing with heavy

thymidine in serum-primed mESCs 2i cultures of R1, cells were passaged twice (5 d) in the same serum-containing medium lacking 2i and simultaneously supplemented with 100 μ M [^{13}C , $^{15}\text{N}_2$]T. For isotope tracing with heavy methionine in serum-primed mESCs, 2i cultures of R1 cells and mESC lines with targeted Tdg alleles were passaged twice (over 5 d) without 2i in L-Met-free DMEM (Life Technologies) supplemented as above and with 0.2 mM of either [methyl- ^{13}C , $^{15}\text{N}_2$]L-Met or natural L-Met. For mESC differentiation without growth factors, R1 cells and the C57Bl/6/129 mixed background cell line were first plated at 1×10^5 cells/cm 2 on gelatin-coated plates in N2B27 medium containing 1,000 U/ml LIF to favor attachment and initial survival⁵¹. After 12 h, the medium was replaced without addition of LIF (defined as time point 0 h). The medium was replaced once more at 24 h. For isotope tracing with [methyl- ^{13}C , $^{15}\text{N}_2$]L-Met during mESC differentiation, R1 cells were cultured for two passages in L-Met-free N2B27 medium supplemented with LIF, 2i and 0.2 mM of either [methyl- ^{13}C , $^{15}\text{N}_2$]L-Met or natural L-Met. Differentiation of R1 cells in the presence of FGF-2 and ActA was as described²⁶, with minor modifications. Briefly, mESCs were cultured in N2B27 medium containing 2i and 1,000 U/ml LIF for several passages and then seeded at 2.2×10^5 cells/cm 2 in N2B27 medium containing 1% KnockOut Serum Replacement (Life Technologies), 12 ng/ml FGF-2 (PeproTech) and 20 ng/ml ActA (ORF Genetics) on plates coated with a thin layer of Geltrex extracellular matrix preparation (Life Technologies). The medium was exchanged after 24 h.

Knockdown experiments in mESCs. shRNA expression vectors targeting Tet1 and Tet2 were generated by cloning synthetic oligonucleotides in pLKO.1 (ref. 52). Recombinant lentiviruses were produced by cotransfecting pLKO.1 shRNA expression vectors and packaging plasmids in HEK-293 cells. 48 h after transduction in the presence of 8 μ g/ml polybrene, shRNA-expressing mESCs were selected with 4 μ g/ml puromycin. Cell pools were continuously cultured in the presence of puromycin. shRNA target sequences were as follows: SCR (control), 5'-CCT AAG GTT AAG TCG CCC TCG-3' (ref. 52); Tet1, 5'-TGT AGA CCA TCA CTG TTC GAC-3' (see ref. 52); Tet2, 5'-TTC GGA GGA GAA GGG TCA TAA-3'. esiRNAs for Smug1 knockdown were generated as described⁵³. The cDNA template for *in vitro* transcription was generated by PCR using following primers: forward, 5'-CGT AAT ACG ACT CAC TAT AGG GAG CCC GTG GGT G-3', and reverse, 5'-CGT AAT ACG ACT CAC TAT AGG GGT TTC GTC CAC TGG G-3'. R1 mESCs were weaned from 2i for two passages in FBS- and LIF-containing medium as described above. Upon plating the second passage, the cells were transfected in a p60 plate with 6 μ g of Smug1 esiRNAs (34.5 nM) and 20 μ l of Lipofectamine RNAi MAX (Life Technologies) according to the manufacturer's instructions and were lysed 72 h after transfection.

Culture and transfection procedures for HEK-293T cells. All transfections were performed using jetPRIME transfection reagent (PEQLAB Biotechnologie GmbH) according to the manufacturer's instructions. HEK-293T cells were seeded 24 h before transfection at a density of 2.5×10^6 cells per 75 cm 2 flask and incubated in 10 ml of medium. The transfection solution (500 μ l of jetPRIME buffer, a specific amount of plasmid DNA (Supplementary Note 4) and 20 μ l of jetPRIME reagent) was added to the medium, and the cells were incubated for 48 h, with an additional medium exchange 24 h after transfection. When cotransfection of esiRNA was performed, a second transfection step (500 μ l of jetPRIME buffer, 5 μ g of esiRNA and 20 μ l of jetPRIME reagent) was carried out 4 h after transfection of plasmid DNA. esiRNAs were purchased from Sigma (human TDG esiRNA EHU038971; human SMUG1 esiRNA EHU098861; human CDK5RAP1 esiRNA EHU079221). **Supplementary Note 4** summarizes the overexpression and knockdown procedures.

Isotope tracing with [^{13}C , $^{15}\text{N}_2$]thymidine or [^{13}C , $^{15}\text{N}_2$]L-Met in HEK-293T cells transfected with Tet1cd. 24 h before transfection, 2.5×10^6 cells were seeded in a 75-cm 2 flask containing 10 ml either of (for [^{13}C , $^{15}\text{N}_2$]thymidine) DMEM medium supplemented with 50 μ M [^{13}C , $^{15}\text{N}_2$]T or (for [^{13}C , $^{15}\text{N}_2$]L-Met) DMEM medium lacking L-Met, L-cystine and pyruvate, which was supplemented with 10% dialyzed FBS, 2 mM [^{13}C , $^{15}\text{N}_2$]L-Met and 0.2 mM L-cystine. Transfection was performed as described above using labeled medium.

Real-time PCR analysis of mRNA expression. For analysis of Tet1 and Tet2 knockdown in mESC total RNA was prepared with Trizol (Invitrogen), cDNA synthesis was performed with Quantitect reverse transcription kit from Qiagen, and real-time PCR was performed with the Power Sybr Green PCR master mix from Applied Biosystems on an Applied Biosystems 7500 Fast

system. Knockdown efficiencies relative to control samples transfected with SCR esiRNAs were 79% and 70% for Tet1 and Tet2, respectively. The primers used to estimate them are listed in **Supplementary Note 4**. For analysis of Smug1 knockdown in mESCs and EpiLC differentiation samples, total RNA was prepared with RNeasy spin columns (Qiagen), followed by DNase treatment using TURBO DNA-free (Ambion, Life Technologies); cDNA synthesis was carried out using iScript cDNA Synthesis kit (Bio-Rad); real-time PCR was performed with SsoFast EvaGreen Supermix (Bio-Rad). Smug1 knockdown efficiency relative to control samples transfected with esiRNAs targeting GFP was estimated to be 60%. Quantification of Tet, Dnmt, Tdg and Smug1 transcripts during EpiLC differentiation and Smug1 knockdown samples was performed using the primers listed in **Supplementary Note 4**. Expression levels were quantified with respect to the housekeeping gene *Gapdh* and normalized to time point 0 h.

Tet *in vitro* assay. A plasmid was prepared from *dam*⁻/*dcm*⁻ competent *E. coli* strain (New England BioLabs) and methylated with M.SssI (New England BioLabs). 1.5 μ g of plasmid DNA were then treated with recombinant Tet1 from the 5hmC TAB-Seq Kit (Wisegene) corresponding to ref. 54. After 3 h incubation at 37 °C and proteinase K treatment, the oxidized plasmids were purified with GeneJET PCR Purification Kit from Thermo Scientific and eluted in 25 μ l water. Samples were then subjected to LC/MS/MS analysis as described⁴². The results are compiled in **Supplementary Note 3**.

Correlation and cluster analyses. Statistical data analysis was performed using IBM SPSS Statistics 19. Results of bivariate correlation analyses are summarized in **Supplementary Note 5**. Unsupervised clustering of species with respect to its correlation coefficients was applied by average linkage hierarchical clustering using a squared Euclidean distance measure.

Pulldown assay. For the pulldown assay, 250 μ g (50 μ l) of the crude nuclear protein extracts were filled up to 500 μ l with 50 mM TEAB and 1 mM MgCl $_2$. The binding conditions were 45 mM TEAB, 1.1 mM MgCl $_2$, 2 mM HEPES, 42 mM NaCl and 20 μ M EDTA. Complete Protease Inhibitor Cocktail Tablets were used from Roche Diagnostics (Indianapolis, IN, USA). The DNA oligomers with DNA-protein cross-linker were dissolved in neat DMSO, and 1 nmol was added to the protein lysate and incubated for 20 min at room temperature. Streptavidin-coated magnetic particles (Roche Diagnostics, Indianapolis, IN, USA) were washed three times with binding buffer (100 mM NaCl, 10 mM Tris, 1 mM EDTA, pH 7.4) before 200 μ l of the bead slurry (equal to 2 mg beads) were added to the sample. Following 2 h incubation at room temperature under constant rotating, the beads were washed three times with 50 mM TEAB and 1 mM MgCl $_2$. The beads were reconstituted in 50 mM TEAB and 1 mM MgCl $_2$. Disulfide bonds of the cross linker were cleaved and alkylated in the process of enzymatic digestion, and the magnetic particles were removed before adding trypsin (described below).

Protein sample preparation. Cell lysis of mouse embryonic stem cells was performed as described in ref. 29. For each lysis, approximately 7.5×10^7 cells were used. Protein concentration was determined by Bradford assay. For each experiment 250 μ g (50 μ l) of the crude nuclear protein extract were used. Protein samples for MS analysis were reduced by adding 100 mM TCEP and by incubating on a shaker at 650 r.p.m. for 1 h at 60 °C and subsequently alkylated with 200 mM iodoacetamide in the dark for 30 min at 25 °C. Protein samples were digested with 0.5 μ g trypsin (Promega, Madison, MA, USA) for 16 h at 37 °C. The reaction was stopped using 1 mM phenylmethylsulfonylfluoride. After tryptic digestion, peptide labeling with the TMT2plex reagents (Thermo Fisher Scientific, Waltham, MA, USA) was performed according to the manufacturer's instructions. TMT2plex reagents 126 and 127 were used to label the samples. When the sample (proteins enriched with hmU- or hmC-containing DNA strands) was labeled with TMT126, the control sample (proteins enriched with no modified DNA-bases) was labeled with the TMT127 reagent and vice versa. Subsequent to the labeling, both sample and control, were combined. This way, each experiment was performed twice as a so-called label swap experiment. Organic solvent was removed by vacuum centrifugation, and the sample was finally reconstituted in 1% (v/v) formic acid for MS analysis.

LC/MS analysis of protein samples. The samples were analyzed using an UltiMate 3000 nano liquid chromatography system (Dionex, Fisher Scientific, Waltham, MA, USA) coupled to an LTQ-Orbitrap XL (Thermo Fisher

Scientific, Waltham, MA, USA). Of each eluate, 15 μ l were injected for the analysis. The samples were desalted and concentrated on a μ -precolumn cartridge (PepMap100, C18, 5 μ M, 100 \AA , size 300 μ m i.d. x 5 mm) and further processed on a custom-made analytical column (ReproSil-Pur, C18, 3 μ M, 120 \AA , packed into a 75 μ m i.d. x 150 mm and 8 μ m picotip emitter). A 57-min multistep analytical separation was performed at a flow rate of 300 nl/min. In the first 50 min, a linear gradient was ramped up from 5% (v/v) solvent B (acetonitrile containing 0.1% formic acid and 5% DMSO) and 95% solvent A (water containing 0.1% formic acid and 5% DMSO) to 95% solvent B. This level was held for 5 min and then ramped down again to 5% solvent B within 2 min. Mass spectrometric analyses were performed starting with a full mass scan in the mass range between m/z 300 and m/z 1,650. This survey scan was followed by three MS/MS scans using the FTMS mass analyzer and high normalized collision energy of 70 in the HCD cell and three additional scans using the ion trap mass analyzer and a normalized collision energy of 35.

Protein identification and relative quantification method. The Thermo Proteome Discoverer 1.1 software (Thermo Fisher Scientific, Waltham, MA, USA) was used for protein identification and for relative quantification. The Sequest (Thermo Fisher Scientific, Waltham, MA, USA) search engine was used in combination with a Uniprot database (*Mus musculus*; date of download, 04/2013). As a limit of detection, a ratio of threefold signal over the noise filter was applied. A maximum of two missed cleavage sites was allowed. The mass tolerances were 10 p.p.m. for the precursor mass and 0.5 Da for the fragment ion mass. Carbamidocysteine was set as static modification. Dynamic modifications were: cation, Na (D, E); the residue of the DNA-protein crosslinker (+146.028 Da; K, Y); Oxidation (M) as well as TMT2plex (N-term. and K). Identified, nonredundant peptides, which were labeled with the TMT2 reagent, were used for relative quantification. The integration window tolerance was 20 p.p.m., and the integration method was set to 'most confident centroid'. The signals of the TMT2 reporter ions 126 and 127 were used to calculate ratios and monitor either preferred or nonpreferred binding of the identified proteins to the modified DNA bases in comparison to the control strand. From the identified

proteins, the only proteins considered as 'specific readers' were enriched in both the forward and the reverse experiment.

42. Schiesser, S. *et al.* Deamination, oxidation, and C–C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine. *J. Am. Chem. Soc.* **135**, 14593–14599 (2013).
43. Gierlich, J. *et al.* Click chemistry as a reliable method for the high-density postsynthetic functionalization of alkyne-modified DNA. *Org. Lett.* **8**, 3639–3642 (2006).
44. Cao, H. & Wang, Y. Collisionally activated dissociation of protonated 2'-deoxycytidine, 2'-deoxyuridine, and their oxidatively damaged derivatives. *J. Am. Soc. Mass Spectrom.* **17**, 1335–1341 (2006).
45. Wang, J. *et al.* Quantification of oxidative DNA lesions in tissues of Long-Evans Cinnamon rats by capillary high-performance liquid chromatography-tandem mass spectrometry coupled with stable isotope-dilution method. *Anal. Chem.* **83**, 2201–2209 (2011).
46. Chen, T., Ueda, Y., Dodge, J.E., Wang, Z. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* **23**, 5594–5605 (2003).
47. Cortazar, D. *et al.* Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**, 419–423 (2011).
48. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J.C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **90**, 8424–8428 (1993).
49. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
50. Montanez, E. *et al.* Kindlin-2 controls bidirectional signaling of integrins. *Genes Dev.* **22**, 1325–1330 (2008).
51. Ying, Q.L. & Smith, A.G. Defined conditions for neural commitment and differentiation. *Methods Enzymol.* **365**, 327–341 (2003).
52. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
53. Kittler, R., Heninger, A.K., Franke, K., Habermann, B. & Buchholz, F. Production of endoribonuclease-prepared short interfering RNAs for gene silencing in mammalian cells. *Nat. Methods* **2**, 779–784 (2005).
54. Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012).

4.10 “Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives”

C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. T. C. Jansen, C. Bauer, M. Münzel, **M. Wagner**, M. Müller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, *Cell* **2013**, 152, 1146-1159.

In diesem Projekt sollte in enger Zusammenarbeit mit der Arbeitsgruppe Vermeulen ein erstes großes Screening nach potentiellen Readern von hmC, fC und caC, d.h. nach mit diesen Basen interagierenden Proteinen durchgeführt werden. Diese Reader sollten durch eine massenspektrometrie-basierte quantitative Proteomanalyse nach der SILAC-Methode identifiziert werden. Als *Bait* sollten dabei modifizierte Oligonukleotide dienen. Für dieses Projekt wurden murine embryonale Stammzellen (ESCs), neurale Vorläuferzellen (NPCs) und Hirngewebe der Maus für eine Analyse ausgewählt.

Wir konnten in dieser Publikation sowohl für mC, als auch für die höher oxidierten Cytosinderivate eine ganze Reihe von potentiellen Readern identifizieren. Die Reader-Sets der einzelnen Modifikationen überlappen sich dabei nur teilweise, so dass für jede Modifikation potentielle spezifische Reader ermittelt werden konnten. Weiterhin konnten wir zeigen, dass das Interaktom der Modifikationen dynamisch reguliert ist. So wurden in ESCs, NPCs und adultem Hirngewebe der Maus jeweils spezifische, nur in einem bestimmten Entwicklungsstadium auftretende Reader der verschiedenen Modifikationen gefunden.

Das Protein UHRF2 wurde im Rahmen dieser Arbeit in neuronalen Progenitorzellen als potentieller hmC-spezifischer Reader identifiziert. Um diesen Befund zu validieren, habe ich für diese Publikation in HEK-Zellen eine Cotransfektion von Uhrf2 und der katalytischen Domäne von Tet1 (Tet1cd), als auch die entsprechenden Kontrolltransfektionen durchgeführt. Aus diesen Proben habe ich die DNA isoliert und die in der DNA enthaltenen Modifikationen *via* LC-MS/MS quantifiziert. Ich konnte zeigen, dass UHRF2 tatsächlich die katalytischen Eigenschaften der TET-Enzyme beeinflussen kann. So habe ich nicht nur einen moderaten Anstieg hmC-Levels im Vergleich zu den Kontrollproben beobachtet, sondern darüber auch einen starken Anstieg des fC- und caC-Levels. Meine Ergebnisse sind in *Figure 6* dargestellt.

Dynamic Readers for 5-(Hydroxy)Methylcytosine and Its Oxidized Derivatives

Cornelia G. Spruijt,^{1,9} Felix Gnerlich,^{2,9} Arne H. Smits,¹ Toni Pfaffeneder,² Pascal W.T.C. Jansen,¹ Christina Bauer,³ Martin Münzel,² Mirko Wagner,² Markus Müller,² Fariha Khan,^{4,5} H. Christian Eberl,⁶ Anneloes Mensinga,¹ Arie B. Brinkman,⁷ Konstantin Lephikov,⁸ Udo Müller,³ Jörn Walter,⁸ Rolf Boelens,⁵ Hugo van Ingen,⁵ Heinrich Leonhardt,³ Thomas Carell,^{2,*} and Michiel Vermeulen^{1,*}

¹Department of Molecular Cancer Research, Proteomics and Chromatin Biology, UMC Utrecht, 3584 CG Utrecht, the Netherlands

²Center for Integrated Protein Science at the Fakultät für Chemie und Pharmazie, Ludwig-Maximilians-Universität München, 81377 Munich, Germany

³Center for Integrated Protein Science at the Fakultät für Biologie, Ludwig-Maximilians-Universität München, 82152 Planegg-Martinsried, Germany

⁴University Institute of Biochemistry and Biotechnology, Pir Mehr Ali Shah Arid Agriculture University Rawalpindi, Rawalpindi, Pakistan

⁵NMR Spectroscopy Research Group, Bijvoet Center for Biomolecular Research, Utrecht University, Padualaan 8, 3584 CH Utrecht, the Netherlands

⁶Proteomics and Signal Transduction, Max-Planck-Institut für Biochemie, 82152 Martinsried, Germany

⁷Department of Molecular Biology, Nijmegen Centre for Molecular Life Sciences, Radboud University Nijmegen, 6525 GA Nijmegen, the Netherlands

⁸Genetik/Epigenetik, Universität des Saarlandes, 66123 Saarbrücken, Germany

⁹These authors contributed equally to this work

*Correspondence: thomas.carell@lmu.de (T.C.), m.vermeulen-3@umcutrecht.nl (M.V.)

<http://dx.doi.org/10.1016/j.cell.2013.02.004>

SUMMARY

Tet proteins oxidize 5-methylcytosine (mC) to generate 5-hydroxymethyl (hmC), 5-formyl (fC), and 5-carboxylcytosine (caC). The exact function of these oxidative cytosine bases remains elusive. We applied quantitative mass-spectrometry-based proteomics to identify readers for mC and hmC in mouse embryonic stem cells (mESC), neuronal progenitor cells (NPC), and adult mouse brain tissue. Readers for these modifications are only partially overlapping, and some readers, such as Rfx proteins, display strong specificity. Interactions are dynamic during differentiation, as for example evidenced by the mESC-specific binding of Klf4 to mC and the NPC-specific binding of Uhrf2 to hmC, suggesting specific biological roles for mC and hmC. Oxidized derivatives of mC recruit distinct transcription regulators as well as a large number of DNA repair proteins in mouse ES cells, implicating the DNA damage response as a major player in active DNA demethylation.

INTRODUCTION

Methylation of cytosine residues at carbon atom 5 of the base (mC) represents a major mechanism via which cells can silence genes. Cytosine methylation mostly occurs in a CpG dinucleo-

tide context. However, CpG islands (CGIs), which are characterized by a very high CpG density and are often found in promoter regions of genes, are typically hypomethylated. Methylation of these CGIs results in transcriptional silencing. The molecular mechanisms underlying the association between DNA methylation and repression of transcription have proven difficult to decipher. The classic view is that methylation of DNA results in the recruitment of methyl-CpG-binding proteins (MBPs) that possess transcriptionally repressive enzymatic activities (Defossez and Stancheva, 2011). However, in vivo validation for this model on a genome-wide level is still lacking. In contrast, recent in vivo data have revealed that CXXC-domain-containing proteins specifically bind to nonmethylated cytosines. In this case, hypomethylated CGIs serve as a recruitment signal for CXXC-domain-containing activators that establish a transcriptionally active chromatin state (Thomson et al., 2010).

It was discovered 4 years ago that Tet enzymes convert mC to 5-hydroxymethylcytosine (hmC) (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). This modification is particularly abundant in the brain and in embryonic stem cells but is detectable in all tissues tested (Globisch et al., 2010; Szwagierczak et al., 2010). Tet enzymes can catalyze further oxidation of hmC to 5-formylcytosine (fC) and 5-carboxylcytosine (caC) (He et al., 2011; Ito et al., 2011; Pfaffeneder et al., 2011). fC and caC can subsequently serve as substrates for thymine-DNA glycosylase (Tdg), which eventually results in the generation of a nonmethylated cytosine (He et al., 2011; Maiti and Drohat, 2011). Therefore, this Tet-Tdg pathway represents an active DNA demethylation pathway. It is not clear whether hmC, fC, and caC have additional DNA-demethylation-independent

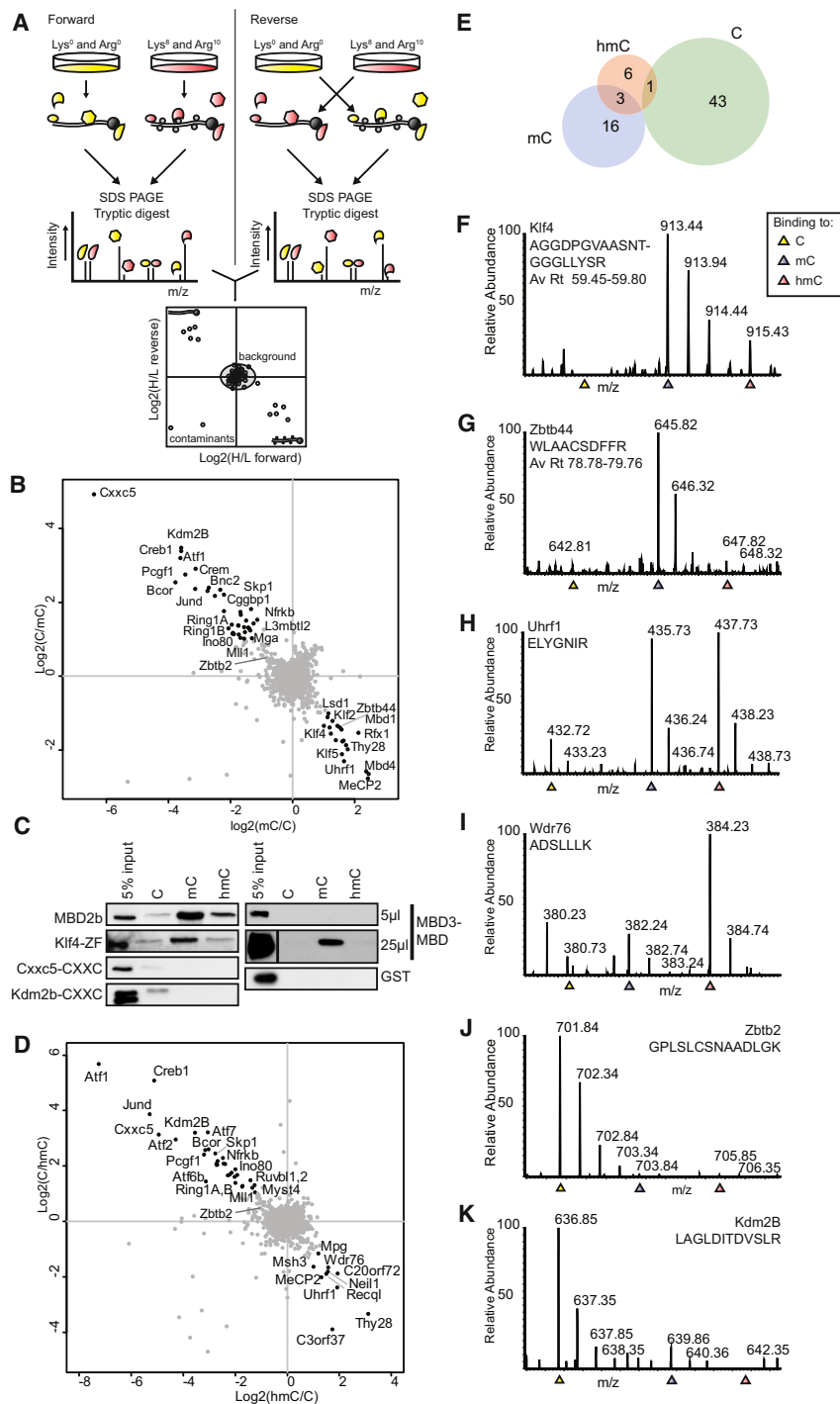


Figure 1. Identification of mC- and hmC-Specific Readers in Mouse Embryonic Stem Cells

(A) Schematic overview of the workflow.

(B) Scatterplot of a SILAC-based mC DNA pull-down in mESC nuclear extracts.

(C) Validation of the mC-specific binding of Klf4 and nonmethyl-C-specific binding of Cxxc5 and Kdm2b. DNA pull-downs were performed with recombinant GST-fusion proteins followed by western blotting. For MBD3_25, an empty lane was cut out.

(D) Scatterplot of a SILAC-based hmC DNA pull-down in mESC nuclear extract.

(E) Venn diagram showing overlap of readers for C, mC, and hmC.

(F–L) Representative mass spectra obtained in the triple-SILAC DNA pull-down in mESCs. Each spectrum shows the relative affinity of the indicated peptides and proteins for nonmethylated (yellow), methylated (blue), and hydroxymethylated (red) DNA.

See also Figure S1 and Table S1.

data reveal that each cytosine modification recruits a distinct and dynamic set of proteins. The known biology of these interacting proteins suggests a role for hmC, fC, and caC in active DNA demethylation pathways via base excision repair (BER), as well as an epigenetic recruitment function in certain cell types.

RESULTS

Identification of mC and hmC Readers in mESCs

To identify readers for methylcytosine and its oxidized derivatives, we made use of a DNA pull-down approach combined with quantitative MS. In brief, nuclear extracts from mESCs grown in “light” or “heavy” SILAC medium were incubated with a nonmodified or modified double-stranded DNA sequence (5′-AAG.ATG.ATG.AXG.AXG.AXG.AXG.ATG.ATG-3′, with X representing C, mC, or hmC; “forward” pull-down; Figure 1A). As a control, a label-swap, or “reverse,” experiment was performed. Following incubation and washes, beads were combined and bound proteins were in-gel digested

with trypsin and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Raw MS data were analyzed using MaxQuant (Cox and Mann, 2008). Specific interactors are distinguishable from background proteins by their H/L ratio. Proteins binding selectively to the modified DNA have a high ratio in the forward pull-down and a low ratio in the reverse pull-down, whereas readers for the nonmodified DNA show opposite

functions, as very few specific binders, or “readers,” for these oxidized versions of mC have been described thus far. We applied quantitative mass spectrometry (MS)-based proteomics to identify a large number of readers for mC and its oxidized derivatives in mouse embryonic stem cells (mESCs). Furthermore, we also identified readers for mC and hmC in neuronal progenitor cells (NPCs) and in adult mouse brain. Our

with trypsin and analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). Raw MS data were analyzed using MaxQuant (Cox and Mann, 2008). Specific interactors are distinguishable from background proteins by their H/L ratio. Proteins binding selectively to the modified DNA have a high ratio in the forward pull-down and a low ratio in the reverse pull-down, whereas readers for the nonmodified DNA show opposite

binding (low forward ratio, high reverse ratio). Background proteins will have a ~1:1 ratio in both pull-downs (Figure 1A).

As shown in Figure 1B and Table S1 available online, we identified 19 proteins enriched for mC compared to C in mESC nuclear extracts ($p < 0.05$ and ratio > 2 in both pull-downs). Among these are the methyl-CpG-binding proteins MeCP2, Mbd1, Mbd4, and Uhrf1 (Defossez and Stancheva, 2011). Other interactors include Rfx1 and Zfx3, which were previously identified as mC readers (Bartke et al., 2010; Sengupta et al., 1999). Interestingly, three Klf proteins were identified as mC readers: Klf2, -4, and -5. These proteins carry three Krüppel-like zinc fingers, just like the Kaiso family of mC-binding proteins. Klf4 is one of the four Yamanaka reprogramming factors and has not been previously identified as a mC-binding protein in HeLa or U937 cells (Bartels et al., 2011; Bartke et al., 2010). This may be due to the low expression of Klf4 in differentiated cells relative to mESCs. We confirmed the direct binding of the Klf4 Krüppel-like zinc fingers to mC using recombinant protein and two different DNA sequences (Figure 1C and S1A). A motif bearing similarities to a recently published consensus binding site for Klf4, as determined by ChIP-seq (GGGXGTG) (Chen et al., 2008), revealed that Klf4 binds this motif with the highest affinity when "X" is mC (Figure S1A). These results establish Klf4 as a sequence-specific mC binding protein.

Mining published bisulfite sequencing data of mESCs and NPCs (Stadler et al., 2012) and overlapping this data with the Klf4 ChIP-seq profile in mESCs (Chen et al., 2008) revealed a substantial number of methylated Klf4-binding sites in this cell type (Figure S1B), which are mainly intronic and intergenic (Figure S1C). Out of the 7,321 Klf4-binding sites in mESCs that were covered in the bisulfite sequencing data set, 1,356 show high levels of DNA methylation in mESCs (18.5%). Many of these Klf4-binding sites contain a methylated Klf4-binding motif, such as GGCGTG (Figures S1D and S1E). Interestingly, many Klf4-binding sites that are nonmethylated in ES cells become hypermethylated in NPC cells (Stadler et al., 2012) (Figures S1B and S1D). This finding may be highly relevant in the context of Klf4-mediated cellular reprogramming. During reprogramming, Klf4 may be able to bind these methylated loci in differentiated cells to initiate stem-cell-specific gene expression patterns. Enrichment analyses for functional domains among the mC interactors revealed DNA-binding zinc fingers to be significantly enriched (Benj.Hoch.FDR = $10^{-2.45}$; Figure S3A). These zinc fingers may also interact with the methylated DNA in a sequence-specific manner.

In addition to the cluster of mC-binding proteins, a large number of proteins displayed preferential binding to nonmethylated DNA (Figure 1B, upper-left quadrant). Consistent with previous observations, this cluster of proteins contains a number of CXXC-domain-containing proteins that are known to preferentially bind to nonmethylated CpGs (Blackledge et al., 2010; Thomson et al., 2010). Examples include Cxxc5, Kdm2b, and Mll1 (see Figure 1C). We also identified other subunits of the Mll1 and PRC1.1 (Bcor) complexes, which most likely bind to the nonmethylated DNA indirectly via Mll1 and Kdm2b, respectively. Other interactors include the Ino80 chromatin-remodeling complex and zinc-finger-containing transcription factors such as Zbtb2, as well as basic leucine zipper-containing proteins

(enriched Benj.Hoch.FDR = $10^{-5.57}$; Figure S3A) such as JunD, Creb1, and Atf7, for which sequence-specific DNA binding is most likely abolished by DNA methylation.

Readers for hmC showed partial overlap with proteins observed to interact with mC (Figure 1D, lower-right quadrant, and Figure 1E), as only three proteins interacted with both modified baits: MeCP2, Uhrf1, and Thy28. Uhrf1 and MeCP2 are known to bind both mC and hmC, although MeCP2 clearly binds with a higher affinity to mC compared to hmC (Frauer et al., 2011; Hashimoto et al., 2012; Mellén et al., 2012). Thy28 is an uncharacterized protein that is associated with apoptosis (Toyota et al., 2012) and contains an EVE domain, which is possibly involved in (ds)RNA binding (Bertonati et al., 2009). Interestingly, two DNA glycosylases (Mpg and Neil3) and a helicase (Recql) were identified as hmC readers in mESCs. These proteins might be involved in active DNA demethylation pathways to convert hmC back to cytosine via base excision repair mechanisms, as has been suggested previously (Hajkova et al., 2010; Wossidlo et al., 2010). In addition, a number of previously uncharacterized proteins, Wdr76 and C3orf37, preferentially bound to hmC compared to C. We purified WDR76 as a GFP fusion protein from HeLa cells and found interactions with OCR, HELLS, and GAN (Figure S1F). The mouse protein Hells, or Lsh, is a DNA helicase that has previously been implicated in regulating DNA methylation levels in cells (Dennis et al., 2001). Interestingly, OCR, or Spindlin-1, is a protein that is known to bind trimethylated H3 lysine 4 (H3K4me3) (Bartke et al., 2010). A large number of proteins preferentially bound to the nonmodified DNA, as was observed for the mC pull-down (Figure S1G). We validated some of these findings using western blotting for endogenous proteins (Figure S1H).

To further investigate the relative affinity of proteins for C versus mC versus hmC in a single experiment, we made use of a triple pull-down approach (Vermeulen et al., 2010), in which mESCs are grown in three different SILAC media. "Light," "medium," and "heavy" nuclear extracts derived from these cells are incubated with C-, mC-, and hmC-containing DNA, respectively (Table S1). Quantitative MS is used to visualize the relative abundance of a protein in each of the three different pull-downs. This experiment confirmed most of the observations made in Figures 1B and 1D, although for some proteins, the ratios in the triple pull-down are lower. As shown in Figures 1F and 1G, Klf4 and Zbtb44 preferentially bind to the methylated DNA. Other proteins bind to both modified baits, such as Uhrf1 (Figure 1H). Kdm2b preferentially binds to the nonmodified DNA (Figure 1K). Contrary to a previous report (Yildirim et al., 2011), we did not observe a specific interaction between MBD3 and hmC (forward ratio, 0.448; reverse ratio, 1.823). We validated these observations using recombinant protein (Figure 1C). At higher concentrations of recombinant MBD3 protein, we observed a specific interaction with mC (Figure 1C), which is in agreement with a recent study that revealed that MBD3 has the highest affinity for mC compared to hmC and C (Hashimoto et al., 2012).

Taken together, these experiments reveal that mC and hmC both recruit distinct proteins in mESCs with little overlap. Furthermore, a large number of proteins preferentially bind to nonmodified DNA. The number of observed interactions with

hmC is moderate, and some of these suggest that hmC acts as an intermediate in active DNA demethylation pathways in mESCs.

fC and caC Recruit a Large Number of Proteins in Mouse Embryonic Stem Cells, Including DNA Glycosylases and Transcription Regulators

We also applied our SILAC-based DNA pull-down approach to identify readers for fC and caC in mESCs. Colloidal blue analysis revealed that the total amount of protein binding to each bait is similar (Figure S2A). Ratios of the forward and reverse pull-downs with hmC, fC, or caC were individually averaged, and these average ratios were then plotted against each other in two-dimensional graphs (Figures 2A–2C and Table S1). From these plots, it is clear that both fC (blue, purple, and green) and caC (yellow and green) recruit many more proteins than hmC does (red and purple). Strikingly, there is only limited overlap between fC and caC binders (green) (Figure 2D). One of the proteins that binds to fC and caC, but not to hmC, is Tdg, which is consistent with its reported substrate specificity (Maiti and Drohat, 2011). We validated this binding behavior using recombinant protein in electrophoretic mobility shift assays (EMSA) (Figures 2E and 2F). We also purified GFP-Tdg from ES cells to identify Tdg interaction partners (Figure S2B and Table S1). None of the Tdg interactors were identified as specific readers in the fC and caC pull-down, indicating that these fC and caC interactions are Tdg independent. Another fC-specific reader is the p53 protein, which plays an important role in DNA damage response (Kastan et al., 1991). Interestingly, Dnmt1 specifically interacted with caC. This interaction was confirmed by EMSA as well as western blotting using an antibody against endogenous protein (Figures 2F and S2C). We also identified subunits of the Swi/Snf chromatin-remodeling complex, such as BAF170, as readers for caC. Three proteins bind to all oxidized derivatives of mC: Thy28, C3orf37, and Neil1. GO term enrichment for biological processes shows that fC significantly enriches for proteins that are related to DNA repair (Benj.Hoch.FDR = $10^{-2.71}$) (Figure S3A), whereas caC interactors are not enriched for any biological process. RNA-binding proteins, mitochondrial proteins, and other proteins that are less likely to be associated with regulation of gene expression or DNA repair were identified as binders for fC and caC (Table S3). Some of these may have a basic affinity for the formyl and carboxyl groups on the DNA strands, which are more reactive than methyl or hydroxymethyl. To exclude the possibility that many fC and caC interactors are binding to damaged or abasic DNA, we validated the homogeneity of the DNA strands using HPLC (Figure S2D). Furthermore, we analyzed the DNA before (blue) and after incubation (red) with mESC nuclear extract by MALDI-TOF-MS (Figure S2E). Quantification of the modified residues by LC-MS/MS shows that there is no significant loss of the modified bases after incubation with nuclear extract (Figure S2F). Figures 2A–2C also show that the group of proteins that bind preferentially to nonmodified cytosine (black, lower-left quadrant) shows a large overlap between the three pull-downs and contains the PRC1.1, Mll1, and Ino80 complexes. To compare the relative affinity of proteins for these three modifications in a single experiment, we performed a triple pull-down. Analyses of the triple pull-down ratios

for the identified fC and caC readers show similar trends, although some of the observed ratios are less prominent. As shown in Figures 2G–2L (and Table S1), the representative spectra of the indicated peptides of Tdg, Neil3, Mpg, Dnmt1, MeCP2, and Uhrf1 show relative ratios that are in agreement with ratios obtained in the independent experiments shown in Figures 2A–2C.

In summary, our data suggest that oxidized cytosine bases may induce a DNA damage response and trigger base excision repair pathways, which may finally result in DNA demethylation. In addition, each of these modifications recruits transcription regulators and other proteins that are not likely to be related to active DNA demethylation.

NPCs Contain a Distinct Set of mC and hmC Readers, Including Uhrf2, which Has the Highest Affinity for hmC

To investigate whether interactions with mC and hmC are dynamic during differentiation, we differentiated mESCs to NPCs. Nuclear extracts were generated from these cells followed by DNA pull-downs. Because no SILAC-compatible neurobasal medium is available, these experiments were performed using label-free quantification (LFQ) (Eberl et al., 2013; Hubner and Mann, 2011). Each DNA pull-down is analyzed separately and in triplicate. For all of the identified proteins (Table S1), we used ANOVA statistics ($p = 0.025$ and $S_0 = 2$) to compare the relative enrichment of proteins for each of the three baits. All significant outliers (192) were hierarchically clustered based on correlation after normalization by row mean subtraction (Figure 3A). Protein enrichment is indicated in red, whereas lack of enrichment is shown in blue. A large number of proteins bind to C or mC, whereas fewer proteins are specifically enriched in the pull-downs with hmC. Three smaller groups of proteins bind specifically to two of the baits (C/hmC, C/mC, or mC/hmC). As was observed in the DNA pull-downs from mESC nuclear extracts, CXXC-domain-containing proteins (Kdm2b and Mll, indicated in black) and their associated factors Bcor/Ring1a/b (blue) and Rbbp5/Ash2l (black) are enriched in the DNA pull-downs with nonmodified DNA relative to mC- and hmC-containing DNA. We identified Mbd2 and associated Mi-2/NuRD complex subunits as mC readers (indicated in yellow). Other identified MBD proteins include Mbd4, MeCP2, and Mbd1. Furthermore, a number of winged-helix (WH)-domain-containing proteins bound specifically to mC, including Rfx5 and its associated factors Rfxap and Rfxank (orange), which have previously been identified as methyl CpG interactors (Bartke et al., 2010).

Strikingly, these proteins bind more strongly to C compared to hmC. We further substantiated these observations by using recombinant protein (Figure 3B). This result indicates that, for some readers, oxidation of mC not only weakens the interaction, but also repels the mC interactor. The homeobox domain is significantly enriched in the cluster of mC-specific readers (Benj.Hoch.FDR = $10^{-1.8}$, Figure S3A), which is consistent with a previous study (Bartke et al., 2010). In addition, several known mC readers, such as Kaiso, Uhrf1, and Mbd4, bind both modified forms of cytosine. A number of DNA glycosylases bind specifically to hmC (Neil1, Neil3), as well as some helicases (Hells, Harp, Recql, and its homolog Bloom), which again suggests a

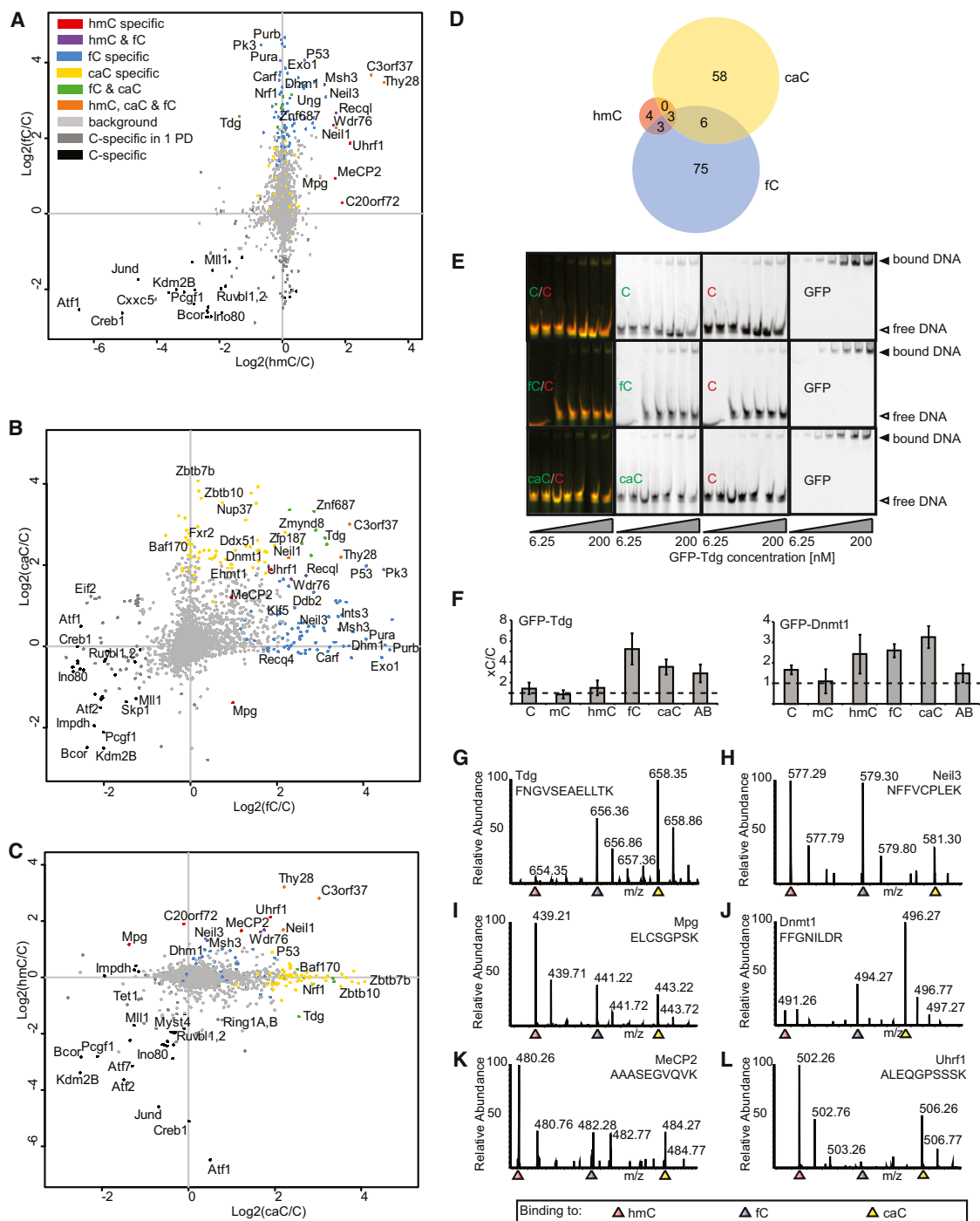


Figure 2. fC and caC Recruit a Large Number of Nonoverlapping Proteins in Mouse Embryonic Stem Cells

(A–C) Scatterplots of SILAC-based mC, fC, and caC DNA pull-downs in mESC nuclear extract. The average ratio of all the identified and quantified proteins in the forward and reverse experiment for each of the three DNA modifications is plotted on the X, Y, and Z axes of a three-dimensional cube. Shown in (A–C) are different side views of the cube. Colors indicate in which of the three pull-downs a protein was significantly enriched.

(D) Venn diagram showing the number of significantly enriched proteins for each of the baits.

(E) EMSA with GFP-Tdg at increasing protein concentrations (6.25–200 nM) incubated with dsDNA (250 nM of differentially labeled xC- and C-containing oligonucleotide, each).

(F) Electrophoretic mobility shift assays as shown in (E) performed with GFP-Tdg and GFP-Dnmt1 for all six residue variants (C, mC, hmC, fC, caC, and abasic site [AB]) in direct comparison to unmodified DNA. The binding preference was determined as the ratio of fluorescence signals of the different DNA substrates in the shifted bands. Shown are the means of three experiments; error bars represent SD.

(legend continued on next page)

DNA-repair-involved DNA demethylation pathway (GO DNA repair; Benj.Hoch.FDR = $10^{-3.91}$; Figure S3A). Although homeobox proteins are known to bind specifically to mC, a number of homeobox proteins show preferential binding to hmC in NPC extracts (examples include Zfx1 and -2). Finally, Uhrf2 was identified as a specific hmC-binding protein in NPCs, which we confirmed using recombinant protein (Figure 3B). Uhrf2 is not expressed in mESCs, and its levels increase upon differentiation (Pichler et al., 2011). This explains why Uhrf2 was not identified as an hmC-specific reader in mESC DNA pull-downs.

Taken together, these experiments reveal that interactions with mC and hmC are highly dynamic during differentiation. Furthermore, the observations made in NPCs strengthen our hypothesis that oxidation of mC serves as a trigger for active DNA demethylation. Nevertheless, some hmC-specific readers in NPCs do not appear to be linked to DNA repair mechanisms, indicating that, in these cells, hmC may also serve a role as a “classical” epigenetic mark that recruits transcriptional regulators.

NMR-Based Analysis of the Rfx5 WH Domain Bound to mC DNA

The specific interaction between the Rfx5 WH domain and mC DNA was studied in detail using solution nuclear magnetic resonance (NMR) spectroscopy in order to derive binding affinity and identify the mC-binding site. Addition of a singly methylated 18bp DNA fragment to the Rfx5-WH domain results in large changes in the ^1H - ^{15}N HSQC “fingerprint” spectrum (Figure 3C). After addition of a slight molar excess of DNA, the spectrum does not show any further changes, indicating that Rfx5-WH strongly binds mC DNA and preferentially at only one of the two mC sites (Figure 3C). The affinity of Rfx5 for mC DNA was derived from the observed peak displacement for residues in the fast exchange regime, such as T104 and E102, assuming that the two mC are independent and equivalent, which resulted in an apparent dissociation constant $K_{D,app}$ of $\sim 3\ \mu\text{M}$ (with 95% probability limits $10\ \text{nM} < K_D < 16\ \mu\text{M}$) (Figure 3D and Supplemental Information). Based on DNA pull-downs done with recombinant protein, which revealed a quantitative depletion of the WH domain from the lysate, we anticipate the K_D to be in the nM range (Figure 3B). To identify the residues that are responsible for specific mC binding, we used the DNA-bound Rfx1 WH domain crystal structure (PDB ID 1DP7; sequence identity 35%; Avvakumov et al., 2008; Gajiwala et al., 2000) to construct a homology model structure of Rfx5-WH and validated it against the experimental chemical shifts (data not shown). The homology model contains a hydrophobic pocket that includes residues with the largest chemical shift changes and is well aligned with an extended basic surface that is responsible for DNA binding in Rfx1. This binding pocket, formed by the side chains of K110, V113, Y114, T132, F135, L139, and Y169, is appropriately shaped to capture the mC base via a flip-out

mechanism, as seen in the case of UHRF1 (Figure 3E). Steric clashes introduced by the presence of an additional hydroxyl group could cause the observed specificity for mC. Given the apparent high affinity and DNA-sequence-independent binding to mC, we propose that the WH domain that is present in Rfx proteins is a bona fide mCpG-binding domain.

Brain-Specific Readers for mC and hmC Include Dlx Proteins

The adult brain is the organ with the highest levels of hmC (Globisch et al., 2010). Tet enzymes and hmC have been shown to play a role in active DNA demethylation of certain genes in this organ (Guo et al., 2011). To identify readers for C, mC, and hmC in the adult brain, nuclear extracts were prepared from this tissue, and these extracts were used for DNA pull-downs. LFQ was used to determine differential binders (Table S1). In brain extracts, we identified fewer specific readers compared to NPCs (108, $p = 0.025$ and $S_0 = 0$; Figure 4), most likely due to the presence of highly abundant structural proteins derived from connective tissue and extracellular matrix in these nuclear extracts. Interestingly, more proteins specifically bind to hmC compared to mC in brain extracts. This is in contrast to NPCs and mESCs, in which more interactions with mC relative to hmC are observed, which may imply a specific role for hmC in brain tissue.

The nonmodified DNA pull-down enriched for the same factors as those observed in mESCs and NPCs, including Cxxc5, Kdm2b, and Bcor (CXXC-domains indicated in black, PRC1 complex in blue, and Ino80 in red). In this case, mC DNA was bound by the Mbd2/NuRD complex, which contains the brain-specific ATPase Chd5 (Eberl et al., 2013; Potts et al., 2011) (indicated in yellow). Interestingly, we identified three distal-less homeobox proteins (Dlx1, -5, and -6) as specific mC interactors. Dlx proteins play a role in the development of the brain and are also expressed in specific regions of the adult brain (Jones et al., 2011; Wang et al., 2011). Wdr76 and Thy28 are hmC specific, as was also observed in NPCs. Thap11 (or Ronin) is identified as a brain-specific hmC reader. Interestingly, this protein is highly expressed in certain regions of the brain, including Purkinje cells (Dejosez et al., 2008). Finally, we identified all four subunits of replication factor C (Rfc2-5) and the associated factor Rfc1 as hmC-specific readers (indicated in green).

Altogether, these experiments further emphasize the dynamic nature of the mC and hmC interactomes during development.

Global Absolute Quantification of Protein Levels in mESCs, NPCs, and Adult Mouse Brain Extracts Reveals Expression-Level-Dependent and -Independent Interaction Dynamics

Our screening for mC- and hmC-specific readers in mESCs, NPCs, and adult mouse brain revealed a large number of cell-type- or organ-specific interactors (Figure S3B). The most

(G–L) Representative spectra of the indicated peptides obtained in the triple-labeled DNA pull-down in mESCs. Each spectrum shows the relative affinity of the indicated peptides and proteins for hmC-containing (red), fC-containing (blue), and caC-containing (yellow) DNA. Spectra are shown for Tdg (G), Neil3 (H), Mpg (I), Dnmt1 (J), MeCP2 (K), and Uhrf1 (L).

See also Figure S2 and Table S1.

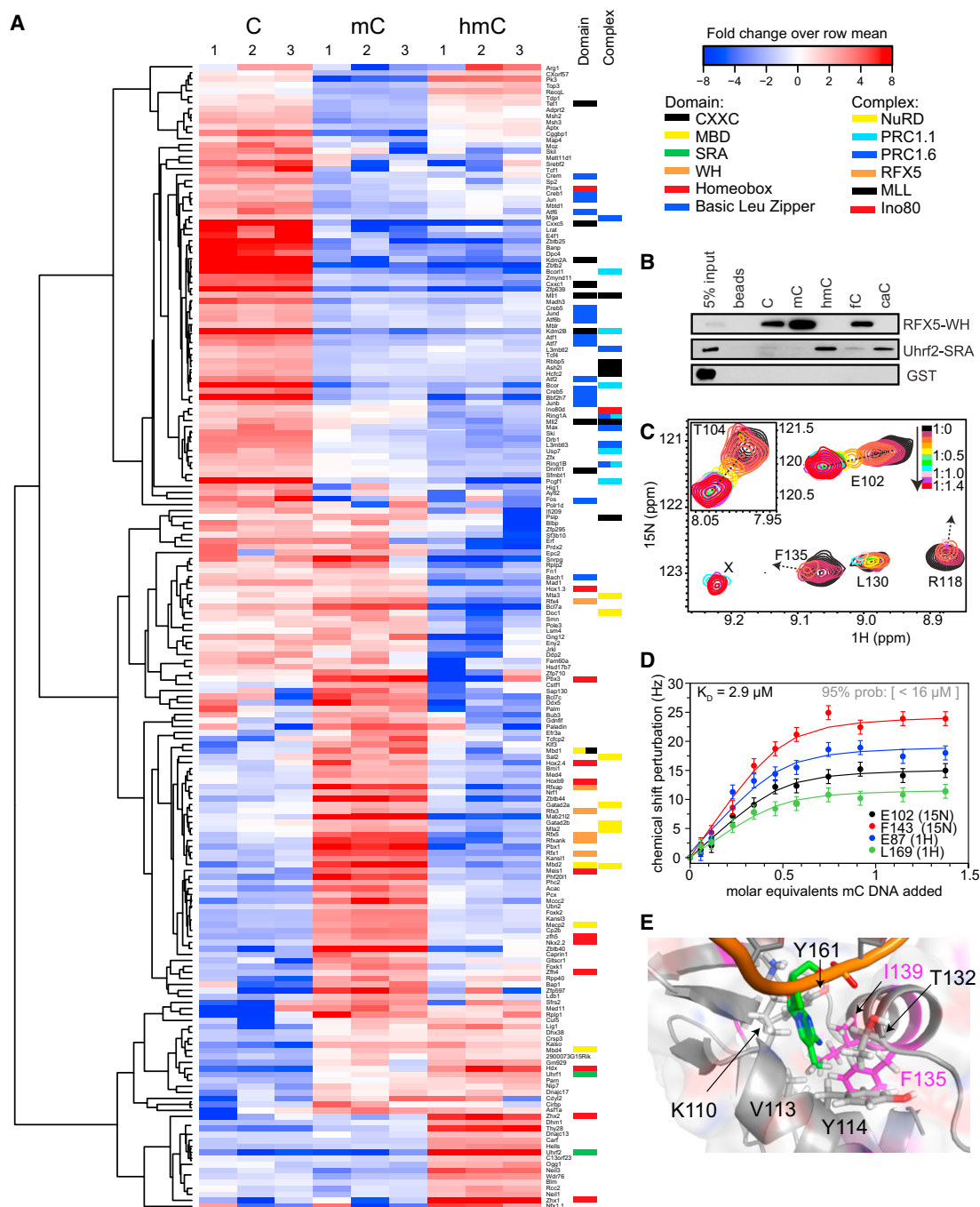


Figure 3. Hierarchical Clustering of NPC-Specific C, mC, and hmC Readers

(A) Correlation-based clustering of the LFQ intensities after log₂ transformation and normalization by row mean subtraction. Included in the clustering are proteins that are significantly binding to at least one of the baits as determined by an ANOVA test. Blue indicates lack of enrichment, whereas enrichment is indicated in red. Domain and Complex columns indicate the DNA-binding domain(s) that may be responsible for direct binding to the bait and the complexes that readers are part of, respectively.

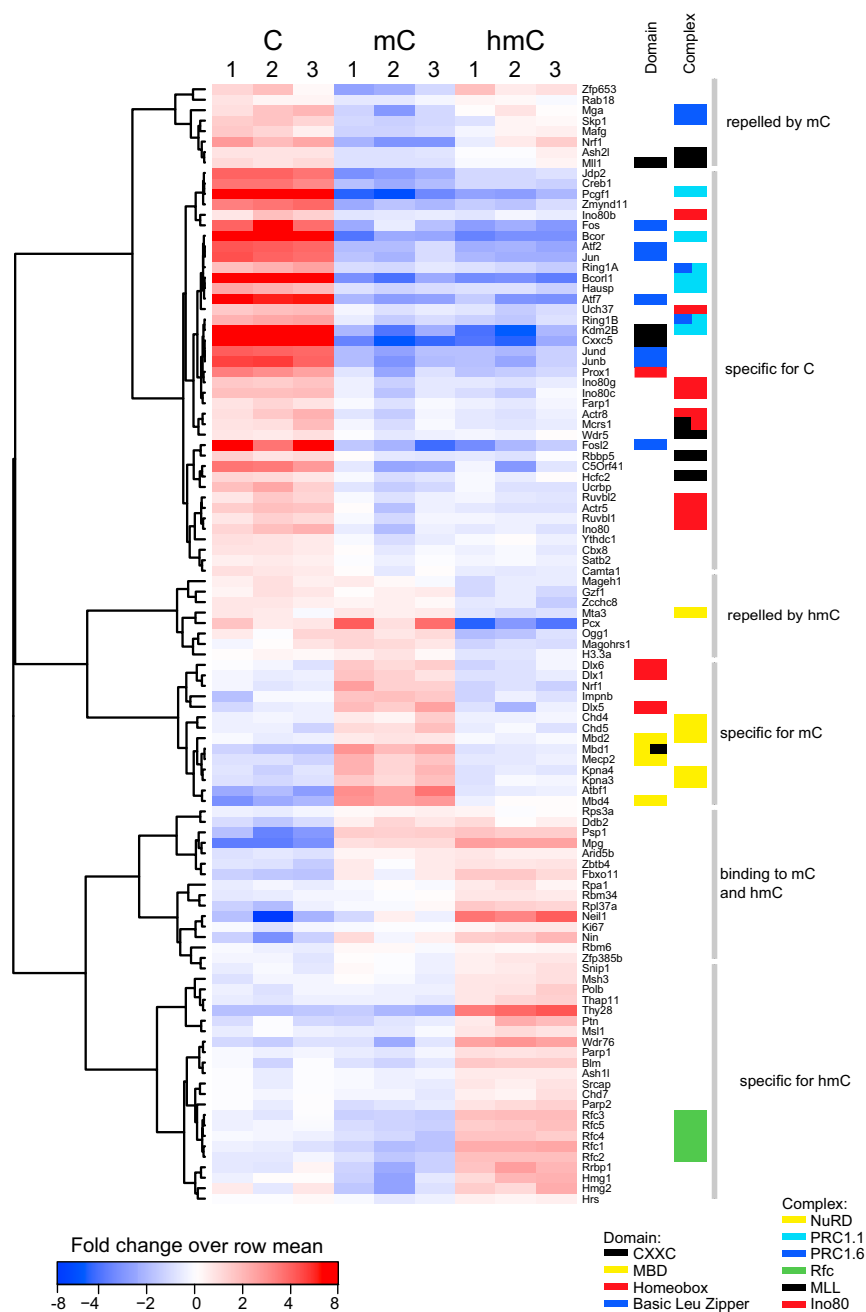
(B) Biochemical validation experiments using DNA pull-downs with recombinant DNA-binding domains.

(C) Overlay of Rfx5-WH HSQC spectra with increasing amounts of mC DNA added and color-coded on the indicated scale listing the WH domain:DNA ratio. Some residues, such as F135 and R118, cannot be unambiguously tracked to their bound states because their chemical shift changes are very large. Peaks corresponding to their bound state, such as "X," appear only after addition of a full molar equivalent of DNA.

(D) Selected binding curves and fits for resonances that are in the fast exchange regime throughout the titration. Error bars (SD) for the peak positions are set to 1.2 Hz.

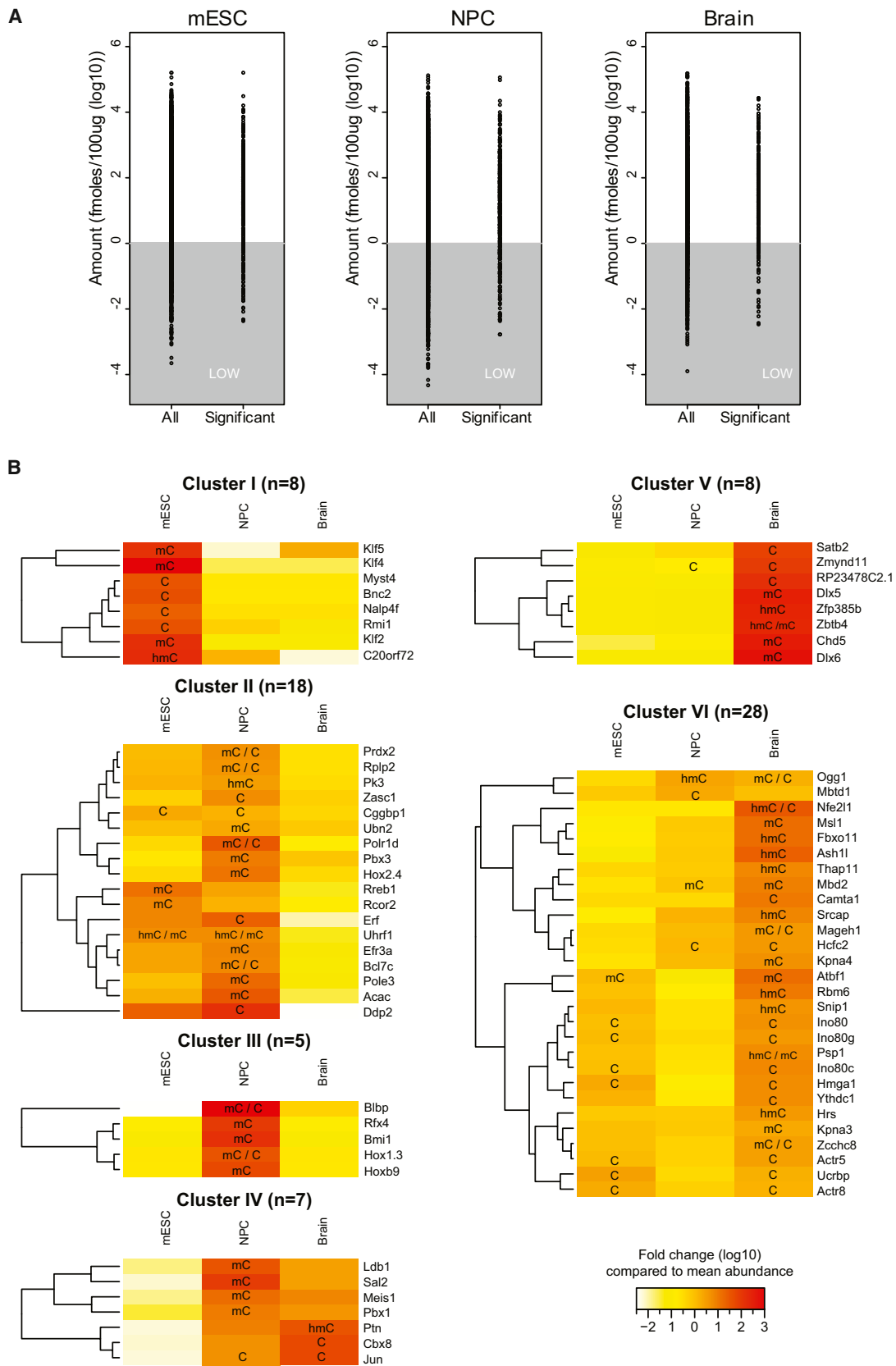
(E) Close-up of the putative mC-binding pocket in the RFX5 WH domain. The methylated cytosine is indicated in green.

See also Table S1.



obvious explanation for these observed differential interactions is regulation of reader abundance at the protein level. Alternatively, the interaction between a reader and (modified) DNA may be affected by posttranslational modifications (PTMs). To investigate global absolute protein levels in the different nuclear extracts that were used for the pull-downs, we made use of a method called intensity-based absolute quantification (iBAQ) (Schwanhäusser et al., 2011). Approximately 8,000 proteins were quantified in at least one of the extracts (Table S2). All proteins with at least a 10-fold change in concentration were clustered based on their expression pattern (Figure S4B).

cause of differential binding may be explained through PTMs that affect the interaction between a reader and DNA or a differentially expressed cofactor. A good example of the latter is the Mi-2/NuRD complex. Although most of its subunits display equal expression levels in mESCs, NPCs, and brain, mC-specific interactions are not observed in mESCs. This can be explained by the fact that Mbd2, which is the direct reader of mC within the NuRD complex, is low in abundance in mESCs and is upregulated during differentiation (Figure 5B). Thereby, it controls the mC-specific binding of the entire complex. In mESCs, the majority of the Mi-2/NuRD complex contains Mbd3, which is



(legend on next page)

the MBD-containing protein that has lost its high-affinity mC binding ability. Furthermore, technical reasons for not identifying an interactor could be the presence of highly abundant structural proteins in the brain lysate or binding competition among different readers in the extracts. Altogether, the absolute quantification of protein abundance in the different nuclear extracts revealed large differences in protein levels between mESC, NPCs, and adult mouse brain. This data set serves as a rich resource on its own but also enables us to explain many of the differential interactions that we identified using quantitative MS-based interactomics.

Uhrf2 Stimulates the Sequential Activity of the Tet1 Enzyme

The first protein that was identified as an hmC binder was Uhrf1 (Frauer et al., 2011), a protein that is involved in maintenance of DNA methylation (Bostick et al., 2007). Our data revealed that Uhrf1 binds with a similar affinity to mC and hmC, which is consistent with previously published data (Frauer et al., 2011). This is in contrast to Uhrf2, which we identified as a high-affinity hmC-binding protein in NPC cells that shows a lower affinity for mC. The function of Uhrf2 is not well understood. It is clear, however, that Uhrf2 cannot rescue the phenotype of Uhrf1 knockout cells, which lose DNA methylation (Pichler et al., 2011; Zhang et al., 2011). Uhrf1 is highly expressed in mESCs, whereas Uhrf2 levels increase during differentiation (Table S3 and Pichler et al., 2011). Altogether, this prompted us to investigate whether Uhrf2 expression affects the levels of mC and its oxidized derivatives. The Tet1-catalytic domain was transfected into HEK293T cells with and without coexpression of Uhrf2. Total genomic DNA modification levels were determined using LC-MS/MS (Figure 6 and Supplementary Information). As shown in Figure 6D, Uhrf2 overexpression increases the level of hmC. More striking is the increase of fC and caC levels upon Uhrf2 co-expression together with the Tet1 catalytic domain. Because fC and caC serve as substrates for Tdg and BER, the detected increase in the levels of fC and caC following Uhrf2 expression may be an underestimation of the actual production of these bases. It therefore seems that Uhrf2 promotes repetitive oxidation of mC by the Tet proteins. We hypothesize that flipping the modified cytosine base out of the DNA double helix, as has been described for Uhrf1 binding to methylated and hydroxymethylated DNA (Arita et al., 2008; Frauer et al., 2011), may enhance accessibility of the hydroxymethylated base to the Tet enzymes, thereby promoting further oxidation.

DISCUSSION

In this study, we have used quantitative MS-based proteomics to identify readers for mC and its oxidized derivatives in mESCs,

as well as readers for mC and hmC in NPCs and adult mouse brain. Readers for individual modifications were found to be highly dynamic throughout the three cell types and tissues that we investigated (Figure 7). This is in contrast to interactions with histone modifications, such as trimethylated lysines on histone H3. For these modifications, the majority of interactors are constant between different cell types or developmental stages (Eberl et al., 2013 and M.V., unpublished data). Readers for distinct cytosine modifications show limited overlap. This indicates that, at least from a biochemical perspective, mC, hmC, fC, and caC behave quite differently. Although little overlap was observed with regard to proteins that interact with each of the epigenetic marks, they all repelled a common set of proteins, such as several CXXC-domain-containing proteins and their interactors. It remains to be determined which of the consequences of DNA (hydroxy)methylation is functionally most relevant: recruitment of transcriptionally repressive complexes or preventing the binding of certain (activating) proteins to unmodified DNA. A detailed biochemical characterization of the interactions and their dissociation constants will be important to answer this question.

Our experiments revealed a number of DNA glycosylases and DNA repair proteins that bind to hmC, fC, and caC, whereas we identified few such proteins binding to mC. The enriched binding of DNA-repair-associated proteins was most pronounced for fC. From this observation, one can conclude that the conversion of hmC to fC is a signal that is likely to result in repair-associated removal of the modified base by proteins that are rather ubiquitously expressed. It is therefore surprising that, in different cell types and tissues, rather constant levels of hmC, fC, and caC are found. The maintenance of such constant levels of these bases in mESCs may indicate a high turnover of DNA methylation, probably involving a constant “correction” by de novo methylation. Regardless, it will be important to investigate which mechanisms control Tet enzyme conversion of mC to hmC and further oxidation to fC and caC. Our data reveal that coexpression of Uhrf2 with the catalytic domain of Tet1 results in a (transient) upregulation of hmC, fC, and caC, indicating that Uhrf2 promotes the sequential oxidation of mC by Tet1. One of the other factors influencing the catalytic activity of the Tet enzymes is the concentration of cellular metabolites. It has been shown that oncometabolites such as 2-hydroxyglutarate can competitively inhibit the activity of 2-oxo-glutarate-dependent enzymes, such as the Tet proteins (Chowdhury et al., 2011; Xu et al., 2011). Furthermore, mutations in IDH1 and -2, which generate 2-oxo-glutarate, are phenocopied by mutations in the TET enzymes and result in cancer (Figuerola et al., 2010). Mutations in the IDH2 and TET2 genes were also linked to lower genomic hmC levels and altered gene expression patterns in myeloid cancers (Ko et al.,

Figure 5. Global Absolute Protein Quantification in mESCs, NPCs, and Adult Mouse Brain

(A) Graphs indicating the concentration of all proteins identified in the nuclear extract (all) and the identified readers (significant) in each of the cell types. The gray area indicates the concentration at which protein quantification is inaccurate.

(B) Readers for which protein expression levels correlate with DNA binding patterns were clustered into six groups based on their expression in the three different nuclear extracts. The color indicates protein levels (white, low; red, high), whereas binding preference is indicated by C, mC, hmC, or combinations thereof.

See also Figure S4 and Table S2 and Table S3.

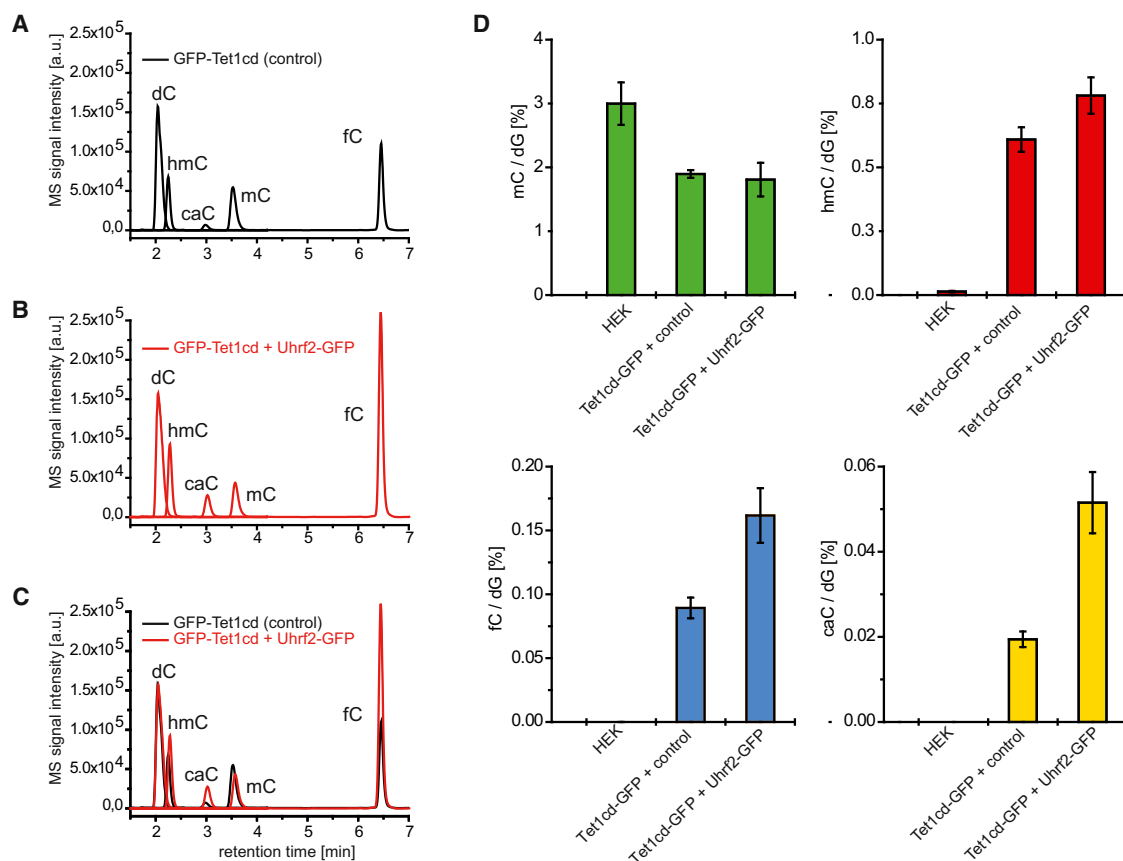


Figure 6. Stable Isotope-Dilution-Based LC-ESI-MS/MS Quantification of Cytosine Derivatives in HEK293T Cells

(A) Nonquantitative LC-MS/MS chromatogram of digested genomic DNA from HEK293T cells cotransfected with Tet1-catalytic domain-GFP (GFP-Tet1cd) and an unrelated expression construct (control). Depicted are the overlaid ion chromatograms of the MS/MS transitions for dC and the cytosine derivatives (black curves). dC, mC, and hmC were measured by a factor of $\sim 10^2$ – 10^3 less sensitive in comparison to caC and fC.

(B) Same as (A) except that Uhrf2-GFP was coexpressed together with GFP-Tet1cd. The MS signal intensities were normalized to the dC content of (A).

(C) Superposition of (A) and (B).

(D) Levels of cytosine derivatives relative to the total cytosine content (dG) as determined by quantitative LC-MS/MS mass spectrometry.

Shown are the means of technical triplicates; error bars reflect SD.

2010; Konstandin et al., 2011). In support of these observations, which clearly link hmC to cancer, we noticed that many hmC, fC, and caC readers are implicated in cancer, including UHRF2, CARF, p53, and HELLS (Lee et al., 2000). Interestingly, mutations in the Hells helicase, which we identified as an hmC reader in NPCs, result in a decrease of DNA methylation levels in cells (Myant et al., 2011). It seems clear that regulating the levels of mC and its oxidized derivatives is essential for normal cell homeostasis and that deregulation of the readers, writers, and erasers of these marks results in a disturbance of the balance between cell proliferation and differentiation during development.

EXPERIMENTAL PROCEDURES

Cell Culture

IB10 mESCs were cultured in light (R^0K^0) or heavy ($R^{10}K^8$) SILAC medium in the presence of 2i compounds. For triple labeling, a third type of medium was used containing medium-labeled L-lysine (K^4) and L-arginine (R^6). mESCs were

differentiated to NPCs in N2B27 medium and cultured in NSA medium, consisting of NSA MEM, 1% glutamine, 1× N2 supplement, 10 ng/ml bFGF, and 10 ng/ml EGF.

DNA Pull-Downs

Nuclear extracts were generated as described previously (Eberl et al., 2013; Vermeulen et al., 2010). DNA (see Table S4) immobilized on Dynabeads MyOne C1 was incubated with nuclear extract in 50 mM Tris-HCl (pH 8), 150 mM NaCl, 1 mM DTT, 0.25% NP40, and complete protease inhibitors (Roche, EDTA-free) in the presence of poly-dAdT. After extensive washes (using incubation buffer without poly-dAdT), bound proteins were in-gel digested using trypsin. After sample preparation, peptides were desalted on Stage-tips (Rappsilber et al., 2003).

Mass Spectrometry

Peptides were separated using an EASY-nLC (Proxeon) connected online to an LTQ-Orbitrap Velos mass spectrometer (Thermo) as described (Smits et al., 2013). Raw data were analyzed using MaxQuant version 1.2.2.5 and searched against protein database iPI.MOUSE.v3.68.fasta. Using Perseus, data were filtered and scatter plots were made using R. The raw mass spectrometry data have been deposited to the ProteomeXchange Consortium

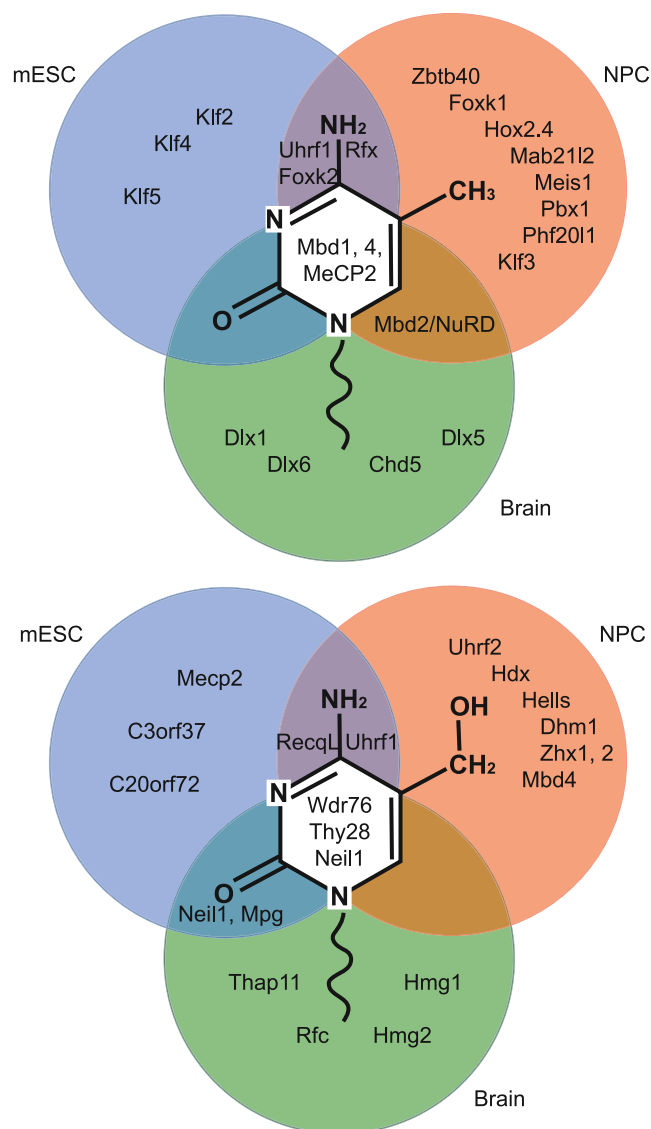


Figure 7. Visualization of Cell-Type-Specific and General mC and hmC Readers

Venn diagram showing examples of mC (A) and hmC (B) readers that were identified in mESCs (blue), NPCs (orange), and adult mouse brain (green). See also Figure S3 and Table S3.

(<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository (Vizcaino et al., 2013) with the dataset identifier PXD000143.

Recombinant Protein Expression and DNA Pull-Downs

DNA-binding domains were cloned into the GST-containing PRP256NB vector. The Uhrf2(aa416–626) GST-fusion construct was kindly provided by Dr. Jiemin Wong. Protein expression was performed in *E. coli* BL21 codon⁺ cells. Bacterial lysate was cleared by ultracentrifugation. DNA pull-downs were performed as described above with the addition of 10 μ M ZnCl₂ to the incubation buffer.

iBAQ

iBAQ was performed essentially as described in Schwanhäusser et al. (2011). A more detailed description can be found in the Extended Experimental Procedures.

LC-MS/MS Analysis of Genomic DNA

Cotransfections were performed in HEK293T cells, and genomic DNA was purified according to Münzel et al. (2010). Quantification of DNA nucleosides from genomic DNA is based on a further development of our isotope dilution method (Pfaffeneder et al., 2011 and data not shown). LC-MS/MS analysis was performed on an Agilent 6490 triple quadrupole mass spectrometer coupled to an Agilent 1290 UHPLC system. For general source- and compound-dependent parameters, see the Extended Experimental Procedures and Tables S5 and S6. The transitions of the nucleosides were analyzed in the positive-ion-selected reaction monitoring mode (SRM) operating MS1 and MS2 under unit mass resolution conditions.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, four figures, and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2013.02.004>.

ACKNOWLEDGMENTS

We would like to thank Dr. Jiemin Wong for providing constructs. We are very thankful to Dr. Nikolay Outchkourov for providing NPCs, and we thank Tony Hyman and Ina Poser for providing the BAC-GFP WDR76 cell line. We would also like to thank Deepani Poramba Liyanage for support. Furthermore, we would like to thank all group members that contributed to fruitful discussions and Susan Kloet for critical reading of the manuscript. We thank the PRIDE team for their help uploading the raw mass spec data. Work in the Vermeulen lab is supported by a grant from the Netherlands Organization for Scientific Research (NWO-VIDI) and an ERC Starting Grant. The Carell group thanks the DFG Normalverfahren CA275/8-4, SFB 749, SFB 646, and the Volkswagen Foundation for financial support. Work in the Leonhardt group is supported by the SFB 646 and SPP 1463 grants from The Deutsche Forschungsgemeinschaft (DFG). M. Münzel and T.P. thank the Fonds der Chemischen Industrie for predoctoral fellowships. F.K. is supported through the International Research Support Initiative Program from the Higher Education Commission, Pakistan. H.v.I. is supported by a Veni fellowship from the Dutch Science Foundation Chemical Sciences, NWO-CW. K.L. and J.W. thank the DFG (WA1029/6) for support through SPP1356.

Received: October 9, 2012

Revised: January 11, 2013

Accepted: February 5, 2013

Published: February 21, 2013

REFERENCES

- Arita, K., Ariyoshi, M., Tochio, H., Nakamura, Y., and Shirakawa, M. (2008). Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. *Nature* 455, 818–821.
- Avvakumov, G.V., Walker, J.R., Xue, S., Li, Y.J., Duan, S.L., Bronner, C., Arrowsmith, C.H., and Dhe-Paganon, S. (2008). Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* 455, 822–825.
- Bartels, S.J.J., Spruijt, C.G., Brinkman, A.B., Jansen, P.W.T.C., Vermeulen, M., and Stunnenberg, H.G. (2011). A SILAC-based screen for Methyl-CpG binding proteins identifies RBP-J as a DNA methylation and sequence-specific binding protein. *PLoS ONE* 6, e25884.
- Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S.C., Mann, M., and Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* 143, 470–484.
- Bertonati, C., Punta, M., Fischer, M., Yachdav, G., Forouhar, F., Zhou, W.H., Kuzin, A.P., Seetharaman, J., Abashidze, M., Ramelot, T.A., et al. (2009). Structural genomics reveals EVE as a new ASCH/PUA-related domain. *Proteins* 75, 760–773.

- Blackledge, N.P., Zhou, J.C., Tolstorukov, M.Y., Farcas, A.M., Park, P.J., and Klose, R.J. (2010). CpG islands recruit a histone H3 lysine 36 demethylase. *Mol. Cell* 38, 179–190.
- Bostick, M., Kim, J.K., Estève, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 317, 1760–1764.
- Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W.W., Jiang, J.M., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* 133, 1106–1117.
- Chowdhury, R., Yeoh, K.K., Tian, Y.M., Hillringhaus, L., Bagg, E.A., Rose, N.R., Leung, I.K.H., Li, X.S., Woon, E.C.Y., Yang, M., et al. (2011). The oncometabolite 2-hydroxyglutarate inhibits histone lysine demethylases. *EMBO Rep.* 12, 463–469.
- Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372.
- Defossez, P.A., and Stancheva, I. (2011). Biological functions of methyl-CpG-binding proteins. *Prog. Mol. Biol. Transl. Sci.* 101, 377–398.
- Dejosez, M., Krumenacker, J.S., Zitur, L.J., Passeri, M., Chu, L.F., Zhou, S.Y., Thomson, J.A., and Zwaka, T.P. (2008). Ronin is essential for embryogenesis and the pluripotency of mouse embryonic stem cells. *Cell* 133, 1162–1174.
- Dennis, K., Fan, T., Geiman, T., Yan, Q.S., and Muegge, K. (2001). Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev.* 15, 2940–2944.
- Eberl, H.C., Spruijt, C.G., Kelstrup, C.D., Vermeulen, M., and Mann, M. (2013). A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics. *Mol. Cell* 49, 368–378. Published online November 29, 2012. <http://dx.doi.org/10.1016/j.molcel.2012.10.026>.
- Figuerola, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H.F., et al. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* 18, 553–567.
- Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M.C., Antes, I., and Leonhardt, H. (2011). Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS ONE* 6, e21306.
- Gajiwala, K.S., Chen, H., Cornille, F., Roques, B.P., Reith, W., Mach, B., and Burley, S.K. (2000). Structure of the winged-helix protein hRFX1 reveals a new mode of DNA binding. *Nature* 403, 916–921.
- Globisch, D., Münzel, M., Müller, M., Michalakakis, S., Wagner, M., Koch, S., Brückl, T., Biel, M., and Carell, T. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* 5, e15367.
- Guo, J.U., Su, Y.J., Zhong, C., Ming, G.L., and Song, H.J. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* 145, 423–434.
- Hajkova, P., Jeffries, S.J., Lee, C., Miller, N., Jackson, S.P., and Surani, M.A. (2010). Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* 329, 78–82.
- Hashimoto, H., Liu, Y.W., Upadhyay, A.K., Chang, Y.Q., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X.D. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* 40, 4841–4849.
- He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q.Y., Ding, J.P., Jia, Y.Y., Chen, Z.C., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303–1307.
- Hubner, N.C., and Mann, M. (2011). Extracting gene function from protein-protein interactions using Quantitative BAC Interactomics (QUBIC). *Methods* 53, 453–459.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300–1303.
- Jones, D.L., Howard, M.A., Stanco, A., Rubenstein, J.L.R., and Baraban, S.C. (2011). Deletion of Dlx1 results in reduced glutamatergic input to hippocampal interneurons. *J. Neurophysiol.* 105, 1984–1991.
- Kastan, M.B., Onyekwere, O., Sidransky, D., Vogelstein, B., and Craig, R.W. (1991). Participation of p53 protein in the cellular response to DNA damage. *Cancer Res.* 51, 6304–6311.
- Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* 468, 839–843.
- Konstandin, N., Bultmann, S., Szwagierczak, A., Dufour, A., Ksienzyk, B., Schneider, F., Herold, T., Mulaw, M., Kakadia, P.M., Schneider, S., et al. (2011). Genomic 5-hydroxymethylcytosine levels correlate with TET2 mutations and a distinct global gene expression pattern in secondary acute myeloid leukemia. *Leukemia* 25, 1649–1652.
- Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930.
- Lee, D.W., Zhang, K.J., Ning, Z.Q., Raabe, E.H., Tintner, S., Wieland, R., Wilkins, B.J., Kim, J.M., Blough, R.I., and Arcenci, R.J. (2000). Proliferation-associated SNF2-like gene (PASG): a SNF2 family member altered in leukemia. *Cancer Res.* 60, 3612–3622.
- Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* 286, 35334–35338.
- Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 Binds to 5hmC Enriched within Active Genes and Accessible Chromatin in the Nervous System. *Cell* 151, 1417–1430.
- Münzel, M., Globisch, D., Brückl, T., Wagner, M., Welzmler, V., Michalakakis, S., Müller, M., Biel, M., and Carell, T. (2010). Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.* 49, 5375–5377.
- Myant, K., Termanis, A., Sundaram, A.Y.M., Boe, T., Li, C., Merusi, C., Burrage, J., de Las Heras, J.I., and Stancheva, I. (2011). LSH and G9a/GLP complex are required for developmentally programmed DNA methylation. *Genome Res.* 21, 83–94.
- Pfaffeneder, T., Hackner, B., Truss, M., Münzel, M., Müller, M., Deiml, C.A., Hagemeyer, C., and Carell, T. (2011). The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* 50, 7008–7012.
- Pichler, G., Wolf, P., Schmidt, C.S., Mellinger, D., Schneider, K., Frauer, C., Fellinger, K., Rottach, A., and Leonhardt, H. (2011). Cooperative DNA and histone binding by Uhrf2 links the two major repressive epigenetic pathways. *J. Cell. Biochem.* 112, 2585–2593.
- Potts, R.C., Zhang, P.S., Wurster, A.L., Precht, P., Mughal, M.R., Wood, W.H., 3rd, Zhang, Y.Q., Becker, K.G., Mattson, M.P., and Pazin, M.J. (2011). CHD5, a brain-specific paralog of Mi2 chromatin remodeling enzymes, regulates expression of neuronal genes. *PLoS ONE* 6, e24515.
- Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* 75, 663–670.
- Schwahnhauser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342.
- Sengupta, P.K., Ehrlich, M., and Smith, B.D. (1999). A methylation-responsive MDBP/RFX site is in the first exon of the collagen alpha2(I) promoter. *J. Biol. Chem.* 274, 36649–36655.
- Smits, A.H., Jansen, P.W., Poser, I., Hyman, A.A., and Vermeulen, M. (2013). Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res.* 41, e28.
- Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F., and Leonhardt, H. (2010). Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* 38, e181.

- Tahiliani, M., Koh, K.P., Shen, Y.H., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935.
- Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W., Deaton, A., Andrews, R., James, K.D., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 464, 1082–1086.
- Toyota, H., Jiang, X.Z., Asakura, H., and Mizuguchi, J. (2012). Thy28 partially prevents apoptosis induction following engagement of membrane immunoglobulin in WEHI-231 B lymphoma cells. *Cell. Mol. Biol. Lett.* 17, 36–48.
- Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G., and Mann, M. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* 142, 967–980.
- Vizcaino, J.A., Côté, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M., Griss, J., Alpi, E., Birim, M., Contell, J., et al. (2013). The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 41(Database issue), D1063–D1069.
- Wang, B., Lufkin, T., and Rubenstein, J.L.R. (2011). Dlx6 regulates molecular properties of the striatum and central nucleus of the amygdala. *J. Comp. Neurol.* 519, 2320–2334.
- Wossidlo, M., Arand, J., Sebastiano, V., Lepikhov, K., Boiani, M., Reinhardt, R., Schöler, H., and Walter, J. (2010). Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. *EMBO J.* 29, 1877–1888.
- Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.H., Ito, S., Yang, C., Wang, P., Xiao, M.T., et al. (2011). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of α -ketoglutarate-dependent dioxygenases. *Cancer Cell* 19, 17–30.
- Yildirim, O., Li, R.W., Hung, J.H., Chen, P.B., Dong, X.J., Ee, L.S., Weng, Z.P., Rando, O.J., and Fazio, T.G. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* 147, 1498–1510.
- Zhang, J.Q., Gao, Q.Q., Li, P.S., Liu, X.L., Jia, Y.H., Wu, W.C., Li, J.W., Dong, S., Koseki, H., and Wong, J.M. (2011). S phase-dependent interaction with DNMT1 dictates the role of UHRF1 but not UHRF2 in DNA methylation maintenance. *Cell Res.* 21, 1723–1739.

4.11 "TET3 is Recruitet by REST for Context-Specific Hydroxymethylation and Induction of Gene Expression"

A. Perera[#], D. Eisen[#], **M. Wagner[#]**, S. K. Laube, A. F. Künzel, S. Koch, J. Steinbacher, E. Schulze, V. Splith, N. Mittermeier, M. Müller, M. Biel, T. Carell, S. Michalakis, *Cell Rep* **2015**, *11*, 283-294.

Co-first authors

In diesem Kooperationsprojekt mit der Arbeitsgruppe Michalakis sollte die biologische Bedeutung der TET-Enzyme und der TET-generierten Basen in der Retina von Mäusen untersucht werden. Die Retina bietet sich in diesem Zusammenhang besonders als Studienobjekt an, weil sie größtenteils aus postmitotischen Neuronen besteht, so dass eventuell auftretende DNA-Demethylierungen aktiv, d.h. enzymatisch erfolgen müssen. Daneben sind zum Zeitpunkt der ersten Augenöffnung der Maus am 12. oder 13. Tag nach der Geburt Veränderungen bei den epigenetisch wirksamen DNA-Modifikationen zu erwarten.

Die Ergebnisse dieses Projektes legen nahe, dass das Enzym TET3 durch das transkriptionshemmende Protein REST / NRSF für eine kontextspezifische Hydroxylierung methylierter DNA-Abschnitte rekrutiert wird. Dies führt in der Retina zu einer Expression zuvor stummgeschalteter Gene.

Für diese Publikation habe ich die massenspektrometrische Quantifizierung der Basen mC, hmC und fC in den Retinaprobe durchgeführt.

TET3 Is Recruited by REST for Context-Specific Hydroxymethylation and Induction of Gene Expression

Arshan Perera,^{1,3} David Eisen,^{2,3} Mirko Wagner,^{2,3} Silvia K. Laube,² Andrea F. Künzel,² Susanne Koch,¹ Jessica Steinbacher,² Elisabeth Schulze,¹ Victoria Splith,¹ Nana Mittermeier,¹ Markus Müller,² Martin Biel,¹ Thomas Carell,^{2,*} and Stylianos Michalakis^{1,*}

¹Center for Integrated Protein Science Munich CiPSM at the Department of Pharmacy – Center for Drug Research, Ludwig-Maximilians-Universität München, 81377 Munich, Germany

²Center for Integrated Protein Science Munich CiPSM at the Department of Chemistry, Ludwig-Maximilians-Universität München, 81377 Munich, Germany

³Co-first author

*Correspondence: thomas.carell@cup.uni-muenchen.de (T.C.), stylianos.michalakis@cup.uni-muenchen.de (S.M.)

<http://dx.doi.org/10.1016/j.celrep.2015.03.020>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

SUMMARY

Ten-eleven translocation hydroxylases (TET1-3) oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC). In neurons, increased 5hmC levels within gene bodies correlate positively with gene expression. The mechanisms controlling TET activity and 5hmC levels are poorly understood. In particular, it is not known how the neuronal TET3 isoform lacking a DNA-binding domain is targeted to the DNA. To identify factors binding to TET3, we screened for proteins that co-precipitate with TET3 from mouse retina and identified the transcriptional repressor REST as a highly enriched TET3-specific interactor. REST was able to enhance TET3 hydroxylase activity after co-expression and overexpression of TET3-activated transcription of REST target genes. Moreover, we found that TET3 also interacts with NSD3 and two other H3K36 methyltransferases and is able to induce H3K36 trimethylation. We propose a mechanism for transcriptional activation in neurons that involves REST-guided targeting of TET3 to the DNA for directed 5hmC generation and NSD3-mediated H3K36 trimethylation.

INTRODUCTION

Epigenetic mechanisms are critically involved in neuronal differentiation and synaptic network formation. 5-hydroxymethylcytosine (5hmC), a recently identified epigenetic DNA modification, is present at high levels in the brain (Münzel et al., 2010) and accumulates during neuronal differentiation (Hahn et al., 2013). 5hmC is generated by ten-eleven translocation (TET) hydroxylases (TET1-3) from 5-methylcytosine (5mC) (Kriacounis and Heintz, 2009; Tahiliani et al., 2009). Accumulation of 5hmC in neuronal genes was shown to correlate positively with gene expression (Colquitt et al., 2013; Hahn et al., 2013; Mellén et al., 2012; Song et al., 2011; Szulwach et al.,

2011), suggesting that 5hmC plays a key role in neuronal differentiation (Santiago et al., 2014). How TET hydroxylase activity is controlled and directed to neuronal genes is not known. The TET1 isoform contains an amino-terminal CXXC domain that is thought to be important for binding to DNA (Xu et al., 2011). Tet2 lacks such a CXXC domain, but was shown to functionally interact with the CXXC domain containing proteins CXXC4 and CXXC5 (Ko et al., 2013; Williams et al., 2011).

TET3, the major TET isoform in neurons (Colquitt et al., 2013; Hahn et al., 2013), exists in various isoforms generated by alternative splicing (Liu et al., 2013), two isoforms containing a CXXC domain and one shorter isoform lacking such a domain. In particular, the TET3 lacking a CXXC domain (neuronal TET3) was shown to be enriched in neuronal tissue (Liu et al., 2013). It is not known how neuronal TET3 is targeted to the DNA. It was suggested that transcriptional regulators might bind directly to TET3, facilitating its specific targeting to genomic positions, and, hence, control 5hmC levels (Xu et al., 2012). Transcription-factor-dependent control of TET3 activity could potentially contribute to epigenetic regulation of gene expression in neurons. However, transcriptional regulators that functionally interact with TET3 have not been identified so far.

To directly address this issue, we investigated the function of TET3 and 5hmC during the formation of a complex neural network: the murine retina. We show that 5hmC levels increase during postnatal retinal network maturation. This gain of 5hmC results in an elevated expression of genes involved in important retinal and/or neuronal functions. To identify factors able to regulate TET3 activity and 5hmC formation in a gene-specific manner, we purified retinal nuclear proteins interacting with TET3. We found that the transcriptional regulator REST specifically interacts with the neuronal TET3 isoform and elevates its hydroxylase activity. Moreover, we show that TET3 overexpression induces the expression of REST target genes. We postulate that transcriptional activation of specific genes results from de novo chromatin remodeling that involves activity of histone 3 lysine 36 methyltransferases like NSD3.

RESULTS

5hmC Is Acquired during Postnatal Retinal Development

Development of neural networks is a complex process that involves the generation, positioning, and synaptic wiring of neurons. This developmental process is finalized with activity-dependent maturation of neurons and refinement of their network. In the mammalian retina, this terminal maturation process involves major morphological and functional changes, largely occurring within a week after eye opening (Hoon et al., 2014; Okawa et al., 2014). These morphological and functional refinements go along with significant changes in the gene expression pattern. To visualize these changes, we applied label-free quantification (LFQ) comparing the retinal protein levels in 2-week-old mice (eye opening) with 3-week-old mice (mature state; Figure 1A), and we found a significant upregulation of proteins involved in synaptic function and visual processing over time (Figures 1A and S1A; Table S1). The molecular mechanisms controlling the underlying transcriptional changes are only partially understood (Gregory-Evans et al., 2013; Xiang, 2013).

We hypothesized that epigenetic regulation at the level of DNA might be involved. First evidence comes from the finding that the levels of the epigenetic DNA modification 5hmC in brain neurons increase with age (Hahn et al., 2013; Münzel et al., 2010). To test for its involvement in terminal retinal maturation, we analyzed the levels of 5hmC during this period. We first used a 5hmC-specific antibody for immunolocalization studies on retinal cryosections and observed a strong increase of the 5hmC immunosignal in all nuclear layers of the retina between week 2 and week 3 (Figures 1B and 1C). At eye opening (week 2), we observed high levels of 5hmC in nuclei of cells within the inner nuclear layer (INL) and the ganglion cell layer (GCL). In contrast, photoreceptor nuclei located in the outer nuclear layer (ONL) were immunonegative (Figure 1B). Strikingly, only 1 week after eye opening (week 3), the levels of 5hmC increased dramatically in both the INL and the GCL. Moreover, rod and cone photoreceptors in the ONL were now positive for 5hmC (Figures 1C and 1D). Confirming the specificity of the anti-5hmC antibody, the immunosignal was quantitatively depleted by 2.5 μ M 5hmC-containing DNA oligonucleotides (Figure 1C, inset). At the sub-nuclear level, most of the 5hmC signal was found to localize in euchromatic regions in both conventional and inverted nuclei (Figures 1E and 1F). In contrast, we found 5mC to be concentrated in pericentromeric heterochromatin (Figures 1E and 1F), suggesting that while 5mC is present in transcriptionally silenced regions, 5hmC is associated with transcriptional activity.

To quantify the levels of 5hmC along with all other cytosine modifications in mouse genomic DNA, we combined ultra-high-performance liquid chromatography with tandem mass spectrometry (UHPLC-MS/MS) (Schiesser et al., 2013; Table S7). The levels of 5hmC increased significantly (Figure 1G and S1B) ($p < 0.005$, Student's *t* test), whereas the levels of the further oxidized derivative 5-formylcytosine (5fC) dropped significantly during the same time frame (Figures 1I and S1C), suggesting distinct roles for 5hmC and 5fC in neurons. Interestingly, the levels of 5mC remained stable during this developmental window (Figures 1H and S1D).

Thus, retinal cells specifically accumulate genomic 5hmC levels during terminal retinal maturation. Importantly, retinal cells

at this developmental stage do not divide; thus, this effect cannot result from increased DNA replication.

To investigate the dynamics of 5hmC and 5mC formation in non-dividing cells, we performed an isotope-tracing experiment by culturing retinal explants from 11-day-old mice for 12 days in the presence of the labeled methyl group donor [$^{13}\text{CD}_3$]-methionine, which provides labeled 5mC (5-[$^{13}\text{CD}_3$]-mC) and 5hmC (5-[$^{13}\text{CD}_2$]-hmC) that can be quantified using UHPLC-MS/MS (Table S8). Confirming the presence of de novo methylation in terminally differentiated retinal cells, we detected 5-[$^{13}\text{CD}_3$]-mC at linearly increasing levels (slope: 0.531, $R^2 = 0.999$) reaching a value of 3.9% of total 5mC after 9 days of [$^{13}\text{CD}_3$]-methionine feeding (Figure 1J). We further detected [$^{13}\text{CD}_2$]-labeled 5hmC (Figure 1K). However, the level of 5-[$^{13}\text{CD}_2$]-hmC was low with only up to 1.6% after 9 days (slope: 0.208, $R^2 = 0.997$). Moreover, the majority of de novo 5hmC was generated from pre-existing and hence unlabeled 5mC (Figure S1E).

5hmC Accumulates in Neuronal Genes and Leads to Increased Gene Expression

To assess the genomic distribution of 5hmC during retinal maturation, we performed antibody-based 5hmC DNA immunoprecipitation (hMeDIP) from 2- and 3-week-old retina followed by next-generation sequencing (NGS). In agreement with the immunohistochemical and UHPLC-MS/MS quantification data, we identified 5,921 intragenic regions that gained 5hmC signal and approximately 20 times less with lost 5hmC signal (Figures 2A and S2A; Table S9). Gene ontology (GO) analysis revealed that genes involved in the morphology of nervous system, CNS development, differentiation of neurons, neuritogenesis, or guidance of axons are significantly enriched within the group of genes that gained 5hmC after eye opening (from week 2 to week 3) (Figure 2B). The genes enriched in the hMeDIP analysis were positively correlated to those enriched in the group of proteins with elevated expression in the mature retina ($p = 6.2 \times 10^{-11}$, Fischer's exact test; Figures 1A and S2B; Table S1). Moreover, 5hmC marks were particularly enriched in gene bodies (Figure 2C), suggesting a regulatory function of 5hmC in gene expression. To investigate if gain of 5hmC is associated with active transcription, we correlated the hMeDIP data with the LFQ protein expression data and found that gain of 5hmC levels from week 2 to week 3 resulted in significantly increased protein expression of the corresponding genes (Figure 2D).

Neuronal TET3 Interacts with Transcriptional Regulators and Histone Writers

5hmC is generated from 5mC by TET hydroxylases. To study if the elevated 5hmC levels in the mature retina result from increased TET-mediated oxidation, we analyzed the expression profile of the three TET enzymes during postnatal development of the mouse retina. We first found that TET3 is the dominant isoform in the retina and its levels increase at eye opening (Figure 3A; Table S10). Importantly, the expression levels at eye opening (postnatal day 13, p13) and at the mature stage (p20) were comparable, ruling out that increased Tet expression causes the increased 5hmC levels (Figure 3A).

TET3 exists in three isoforms that differ in their N-terminal sequence: two isoforms with a CXXC DNA-binding domain and

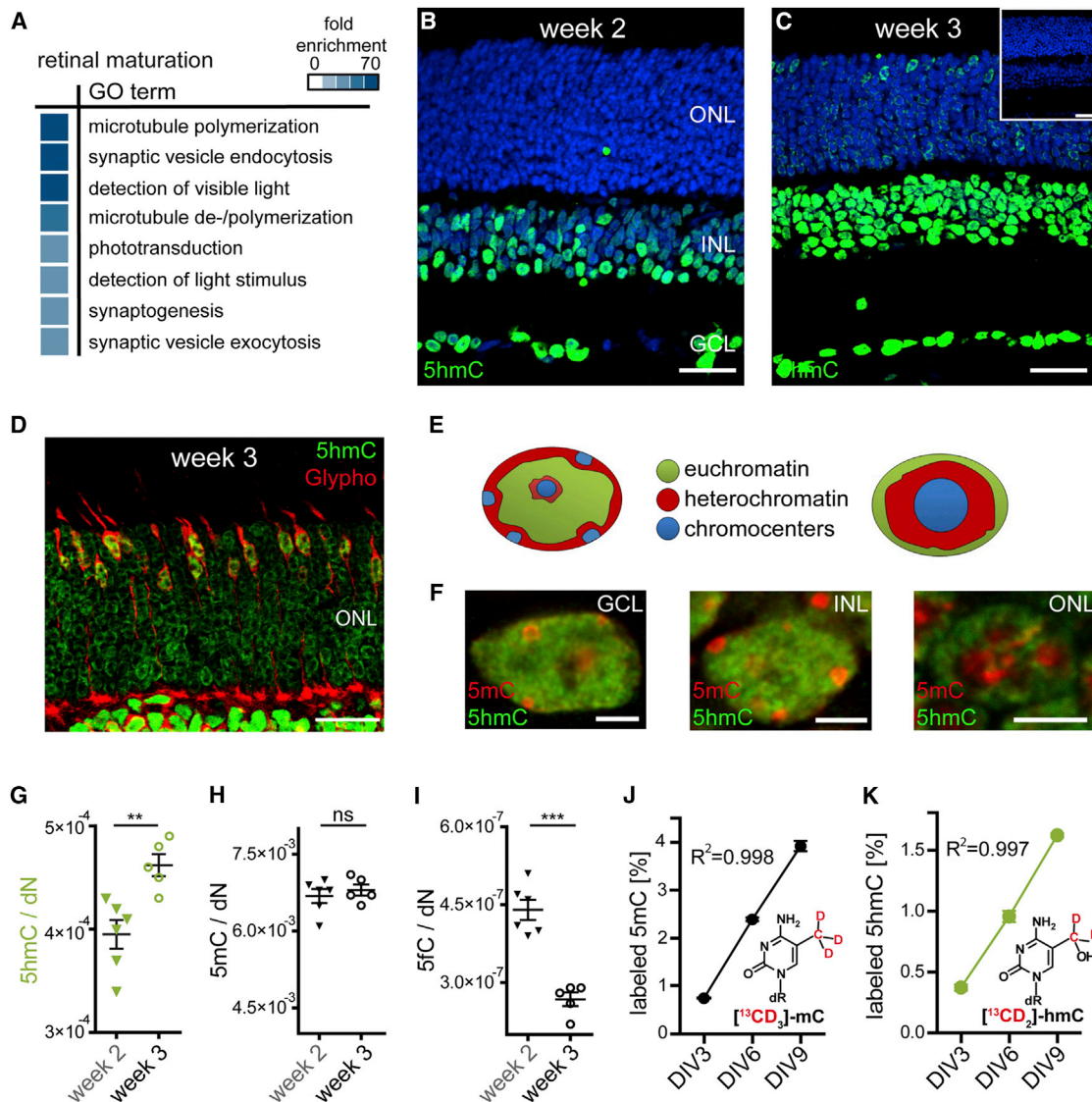


Figure 1. Transcriptional and Epigenetic Changes during Retinal Maturation

(A) GO analysis of upregulated proteins during retinal maturation (from eye opening at week 2 to terminal maturation of the retina at week 3) identified by LFQ is shown.

(B and C) Confocal scans from mouse retinal slices at week 2 (B) and week 3 (C) immunolabeled for 5hmC (green). There is a marked increase in 5hmC signal in nuclei of all retinal layers from week 2 to week 3, in particular in photoreceptors (ONL). The inset in (C) is from a depletion experiment proving the specificity of the anti-5hmC antibody.

(D) Magnification of ONL stained for 5hmC (green) and glycogen phosphorylase (Glypho) (red) is shown.

(E) Schematic view of conventional and inverted nuclei depicts heterochromatin and euchromatin regions.

(F) High-magnification images of conventional nuclei within the GCL or the INL and a rod nucleus with inverted nuclear architecture (Solovei et al., 2009) co-immunolabeled with 5hmC (green) and 5mC (red). In all cases, 5hmC is mainly found in euchromatin whereas 5mC is localized in heterochromatic regions.

(G–I) Scatter plots of UHPLC-MS/MS quantification of global (G) 5hmC, (H) 5mC, and (I) 5fC in the mouse retina at weeks 2 and 3 reveal an age-dependent increase in 5hmC levels, stable 5mC levels, and decreasing 5fC levels.

(J and K) Isotope tracing combined with UHPLC-MS/MS in mouse retinal explant cultures. Retinal explant cultures were fed with [methyl- 13 CD $_3$]-L-methionine and harvested for UHPLC-MS/MS quantification of labeled 5mC (J) and 5hmC (K) after 3, 6, and 9 days in vitro (DIV). Summary data are mean \pm SEM in (G–I) and mean \pm SD in (J and K). **p < 0.01, ***p < 0.001 (Student's t test). The cell nuclei in (A and B) were stained with Hoechst 3442 nuclear dye (blue). Scale bar, 25 μ m in (B–D) and 3 μ m in (F). See also Figure S1 and Tables S7 and S8.

a short isoform enriched in neuronal cells lacking the CXXC domain (Liu et al., 2013). Using isoform-specific primer sets, we found that the short variant lacking the CXXC domain is the

major TET3 isoform in the retina (Figure 3B). Other than the isoforms with the CXXC domain, this neuronal TET3 variant has no DNA-binding affinity (Xu et al., 2012). We hypothesized that

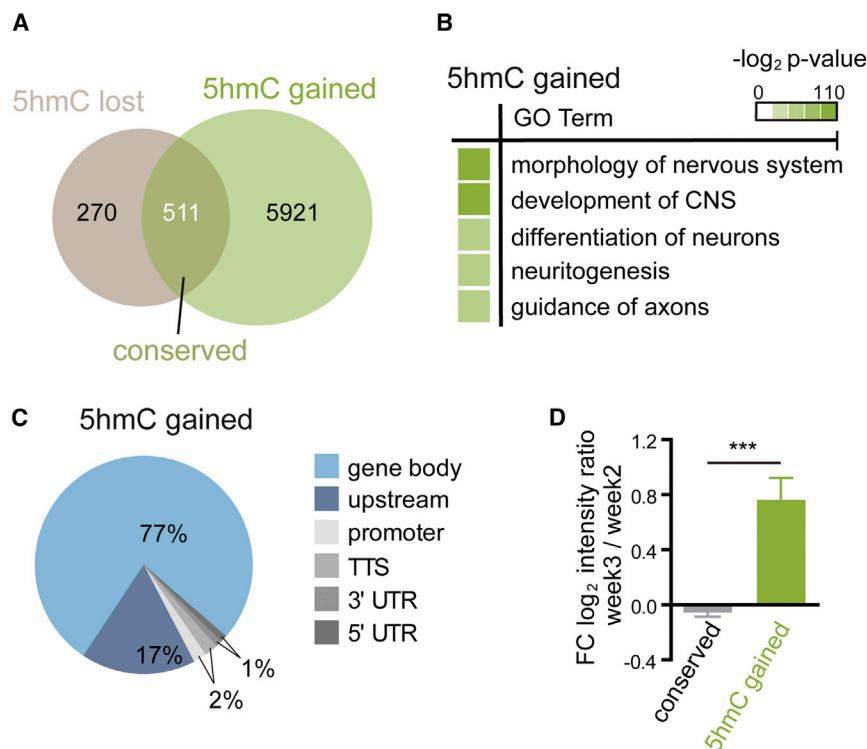


Figure 2. Genomic Localization of 5hmC in the Mouse Retina and Its Effect on Gene Expression

(A) Venn diagram shows 5hmC dynamics during retinal maturation (postnatal week 3/postnatal week 2).
(B) GO analysis of genes that gained 5hmC during retinal maturation is shown.
(C) Pie chart shows intragenic distribution of 5hmC regions in mature retina.
(D) LFQ of proteins encoded by genes with conserved (gray) or gained (green) 5hmC levels during retinal maturation. Summary data are mean \pm SEM. ***p < 0.001 (Student's t test). TSS, transcription starting site; TTS, transcription termination site. See also Figure S2 and Table S9.

unknown DNA-binding (transcription) factors could bind and target neuronal TET3 to pre-existing 5mC for subsequent oxidation to 5hmC (Figure 3C). To identify nuclear proteins that have the potential to specifically bind to neuronal TET3 in mouse retina, we combined affinity purification with mass spectrometry (MS). To this end, we transduced mouse retinal explant cultures with lentiviral vectors expressing an eGFP-fusion protein of the neuronal TET3 splice variant. Control cultures were transduced with lentiviral vectors expressing eGFP only. After onset of expression, nuclear protein complexes were extracted, affinity-purified, and subsequently labeled with TMT 2-plex reagents in forward and reverse label-swap experiments for comparative quantification using LC-MS (see the Experimental Procedures). These experiments allowed us to identify 52 proteins that were significantly enriched for binding to TET3 (>1.5-fold enrichment) in forward and reverse labeling experiments (Figure 3D, bottom right quadrant). Notably, nine of those proteins, such as REST, SUZ12, and CEP290, could be classified as regulators of transcription (Figure 3D; Table S2). Another six proteins are histone variants involved in chromatin modification, i.e., H3F3A and H2AFJ (Figure 3D; Table S2).

TET3 is endogenously expressed in the mouse retina (Figure 3A). Thus, TET3-eGFP is expected to compete for binding partners with endogenous TET3. To identify proteins that bind to endogenous TET3, we repeated the affinity purification experiment using a TET3-specific antibody. Endogenous TET3 was highly enriched by the antibody highlighting its specificity (Figure 3E). Moreover, we identified five transcriptional regulators (ASXL1, CTCF, MORF4L1, REST, and VAX1), three H3K36 methyltransferases (NSD2, NSD3, and SETD2), and histone 3 (H3F3A)

as interactors with significant enrichment for binding to endogenous TET3 (>1.5-fold) (Figure 3E; Table S3). Interestingly, the TET3-interacting protein with the highest enrichment score was the transcriptional repressor REST (Figure 3E), which was also significantly enriched in the TET3-eGFP affinity purification experiment (Figure 3D). To verify the TET3-REST interaction, we performed a reverse experiment using a REST-specific antibody that efficiently immunoprecipitates endogenous REST from mouse retina (Figure 3F; Table S4). Importantly, in this experiment, the REST antibody affinity-purified TET3 very efficiently with the overall third-highest enrichment score, but failed to immunoprecipitate TET1 and TET2 (Figure 3F). Moreover, none of the identified TET3 peptides was specific for the long CXXC-containing isoforms, suggesting that REST interacts specifically with the neuronal TET3 isoform. Taken together, these experiments identify REST as a major specific interacting protein of neuronal TET3.

REST Recruits Neuronal TET3 to Mediate 5hmC Formation and Transcriptional Activation

To determine the functional effects of REST on TET3 hydroxylase activity, we overexpressed TET3-eGFP alone and together with REST in HEK293T cells and subsequently quantified the 5hmC levels (Figure 4A). Non-transfected HEK293T cells have very low endogenous 5hmC levels, and overexpression of TET3-eGFP alone resulted in increased 5hmC levels (Figure 4A). Importantly, however, after co-expression with REST, TET3-eGFP generated significantly higher levels of 5hmC (Figure 4A). Conversely, after small interfering RNA (siRNA)-mediated knock-down of endogenous REST, overexpression of TET3-eGFP led to a significantly weaker elevation of 5hmC levels compared to control siRNA treatment (Figure 4B). Thus, REST is both necessary and sufficient to elevate TET3 hydroxylase activity.

REST is a transcriptional repressor that binds to specific RE-1 binding sites present in regulatory regions of its target genes (Kraemer et al., 1992). Binding of REST results in transcriptional repression of the corresponding genes (Chong et al., 1995; Schoenherr

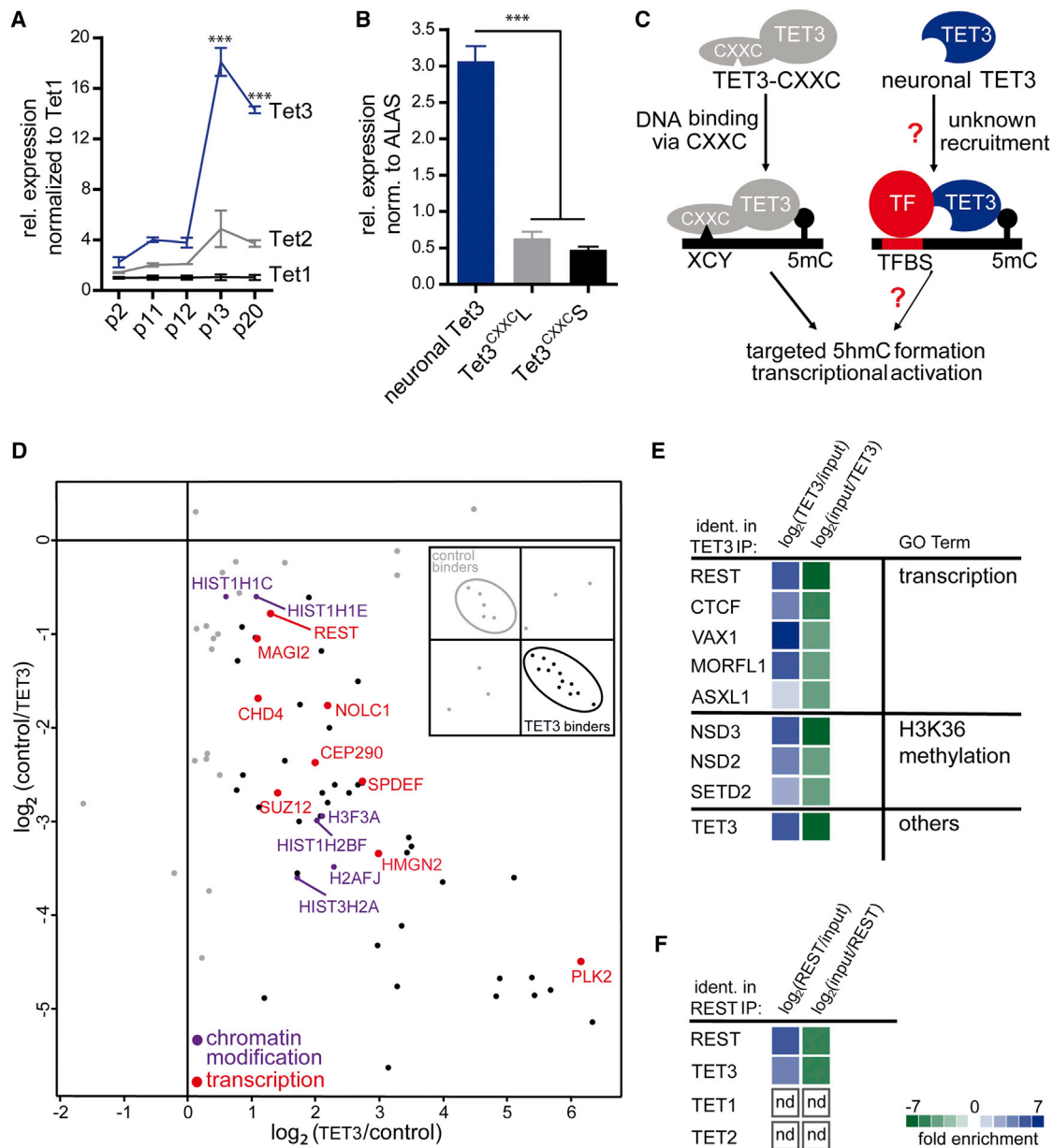


Figure 3. Tet Expression Levels and Identification of TET3-Interacting Proteins in Retina

(A) qRT-PCR analysis of Tet1, Tet2, and Tet3 gene expression in mouse retina at postnatal day (p) 2, 11, 12, 13, and 20 normalized to corresponding Tet1 levels. Tet3 is expressed significantly higher than Tet1 and Tet2 at p13 and p20.

(B) qRT-PCR analysis of Tet3 isoforms. Tet3 lacking a CXXC DNA-binding domain (neuronal Tet3) is the main isoform in retina.

(C) Cartoon illustrating putative DNA-binding mechanisms of TET3 isoforms lacking (neuronal TET3) or containing a CXXC domain. We hypothesized that neuronal TET3 is recruited to the DNA by transcription factors for context-specific hydroxymethylation and induction of gene expression.

(D) Scatter plot is shown of TET3 interactors identified using lentiviral TET3-eGFP overexpression (OE) in retinal explant cultures followed by GFP-trap and LC-MS/MS.

(E) Heatmap of endogenous TET3 interaction partners identified by immunoprecipitation and LC-MS/MS is shown.

(F) Heatmap of endogenous REST immunoprecipitation LC-MS/MS experiment confirms the interaction with endogenous TET3 (nd, not detected). Summary data are mean \pm SEM in (A and B). *** $p < 0.001$ (one-way ANOVA). See also Table S10.

and Anderson, 1995). We therefore asked whether TET3 might affect gene expression of REST target genes. To address this question, we overexpressed TET3-eGFP in HEK293T cells that endogenously contained REST (Dietrich et al., 2012), and we

determined the global protein levels using LFQ (Figures 4C–4E). Overexpression of TET3-eGFP resulted in significant upregulation of 40 proteins, and 26 of those were previously non-expressed (proteins with infinite log2 ratios in Figure 4E).

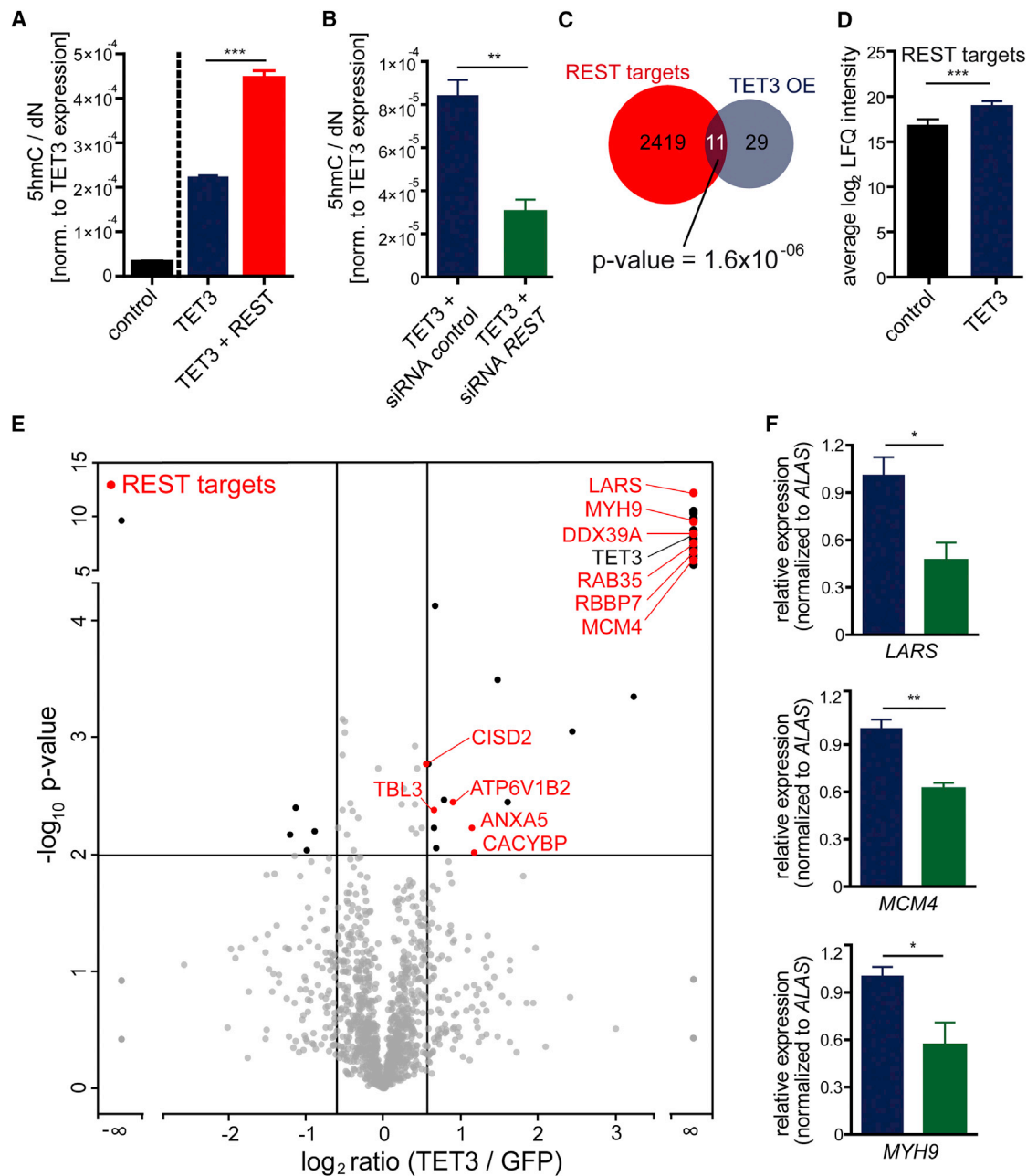


Figure 4. Functional Analysis of the Interaction of Transcriptional Repressor REST with Neuronal TET3

(A) UHPLC-MS/MS quantification of global 5hmC levels after OE of TET3-eGFP or REST and TET3-eGFP is shown.

(B) UHPLC-MS/MS quantification of global 5hmC levels after OE of TET3-eGFP in the presence of siRNA directed against human REST or a scrambled siRNA control (ctrl) is shown.

(C) Identification of global REST target gene expression using LFQ after eGFP (control) or TET3-eGFP OE in HEK293T cells. REST target genes are expressed significantly higher after TET3 OE.

(D) Venn diagram shows the significant overlap (Fisher's exact test; p value = 1.6×10^{-6}) of proteins enriched after TET3 OE and REST target genes in HEK293T cells.

(E) Volcano plot of protein expression ratios between TET3-eGFP and eGFP OE experiments in HEK293T cells as a function of statistical significance (Student's t test p value ≤ 0.01). Proteins with no statistically significant difference in expression between subsets are gray. Proteins with no detectable signal in one of the subsets were assigned a ratio of infinity. REST target genes are highlighted in red.

(F) Quantification of gene expression of three TET3-induced REST target genes in HEK293T cells using qRT-PCR after OE of TET3-eGFP in the presence of siRNA directed against human REST (green) or a scrambled siRNA control (ctrl, blue). Summary data in (A–C and F) are mean \pm SEM. *p < 0.05, **p < 0.005, ***p < 0.001 (Student's t test). See also Table S11.

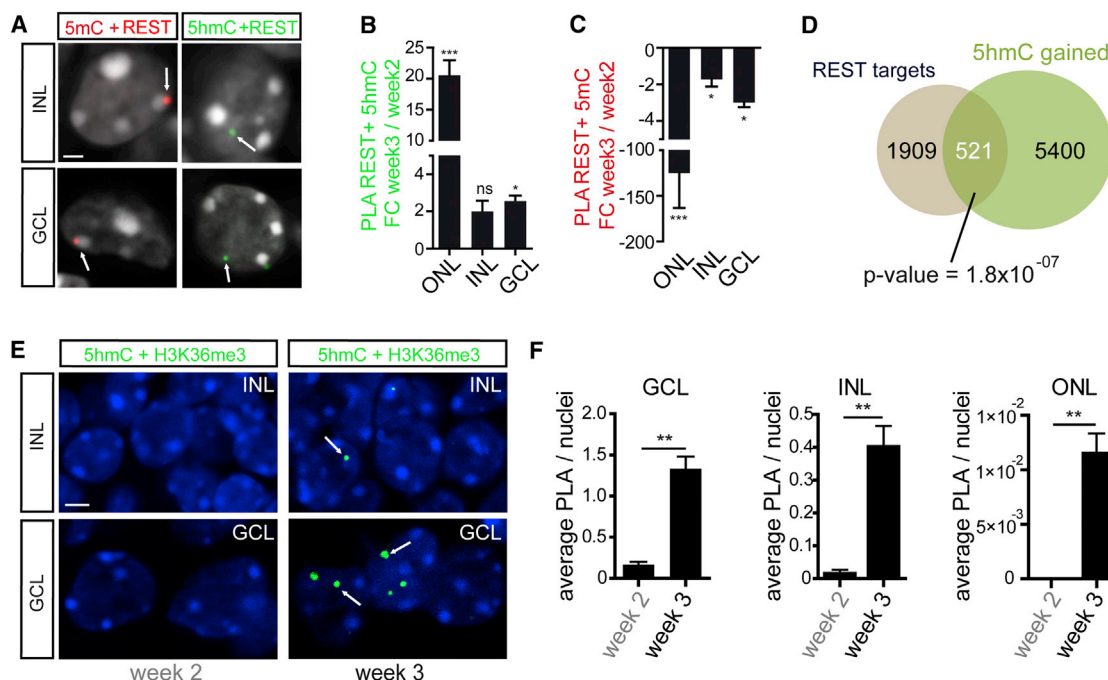


Figure 5. Accumulation of 5hmC in REST Target Genes Is Associated with Chromatin Remodeling

(A) PLA signal for 5mC and REST (red) is localized in heterochromatin. In contrast, the PLA signal for 5hmC and REST is localized in euchromatin (green).

(B and C) Quantification data of the 5hmC/REST (B) and 5mC/REST (C) PLA experiments in (A) are shown.

(D) Venn diagram shows the significant overlap (Fisher's exact test; p value = 1.8×10^{-07}) of REST target genes and 5hmC-gained genes during retinal maturation.

(E) PLA signal for 5hmC and H3K36me3 (green) in nuclei of the INL (top) and the GCL (bottom) at week 2 (left) and week 3 (right).

(F) Quantification data of the 5hmC/H3K36me3 PLA experiments in (E) are shown. Summary data in (B, C, and F) are mean \pm SEM. *p < 0.05, **p < 0.01, ***p < 0.001 (Student's t test). Scale bar, 1 μ m in (A) and 5 μ m in (E).

Interestingly, 308 of the identified proteins were known REST targets (Lu et al., 2014), and the average log₂ LFQ intensity of those REST targets was significantly higher after TET3 overexpression compared to the eGFP control (Figure 4C). Importantly, 11 of the 40 TET3-induced proteins ($p = 1.6 \times 10^{-6}$, Fischer's exact test) were REST targets (Figure 4D; Table S5) and six of those (DDX39A, LARS, MCM4, MYH9, RAB35, and RBBP7) were previously non-expressed proteins (Figure 4E; Table S5).

To test if this TET3-mediated induction of REST target genes in HEK293T cells depends on REST, we overexpressed TET3-eGFP in the presence of REST-specific siRNA or scrambled control siRNA and measured the transcript levels of *LARS*, *MCM4*, and *MYH9* (Table S11). As shown in Figure 4E, all three REST target genes were exclusively expressed in the presence of TET3. Interestingly, in all three cases, the positive effect of TET3 on gene expression was significantly lower after knock-down of endogenous REST (Figure 4F), confirming that TET3 requires the presence of REST to de-repress REST target genes.

Accumulation of 5hmC in REST Target Genes Is Associated with Chromatin Remodeling

These results indicate that REST directs TET3 to repressed target genes and subsequent hydroxylation of pre-existing 5mC, then leads to de-repression and transcriptional activation of the genes. Repressed genes are usually found in hetero-

chromatin, whereas actively transcribed genes are located in euchromatin. As shown in Figure 1F, 5mC is enriched in heterochromatic structures and 5hmC in euchromatin. To test for the co-localization of REST with 5mC and 5hmC in nuclei of retinal neurons, we applied an in situ proximity ligation assay (PLA) combining REST-specific with 5mC- or 5hmC-specific antibodies. PLA gives only a positive signal when 5mC or 5hmC and REST are in close proximity (<40 nm) (Söderberg et al., 2006). We observed a positive PLA signal for both combinations, REST/5mC as well as REST/5hmC (Figure 5A). Importantly, the PLA signal for REST/5mC was localized in heterochromatin (Figure 5A, top), whereas the REST/5hmC signal was observed in euchromatin (Figure 5A, bottom). Thus REST-binding sites show distinct subnuclear localization depending on the oxidation status of adjacent genomic cytosines. Interestingly, the levels of PLA signal for REST/5mC and REST/5hmC varied significantly between week 2 and week 3 (Figures 5B and 5C). In particular, the REST/5hmC PLA signal correlated positively with the 5hmC levels increasing from week 2 to week 3 (Figure 5B). In contrast, the PLA signal for REST/5mC showed an inverse behavior, decreasing during the same period of time (Figure 5C). Together, these results suggest a functional correlation between REST-binding sites and proximal epigenetic DNA marks that might affect chromatin state and gene expression.

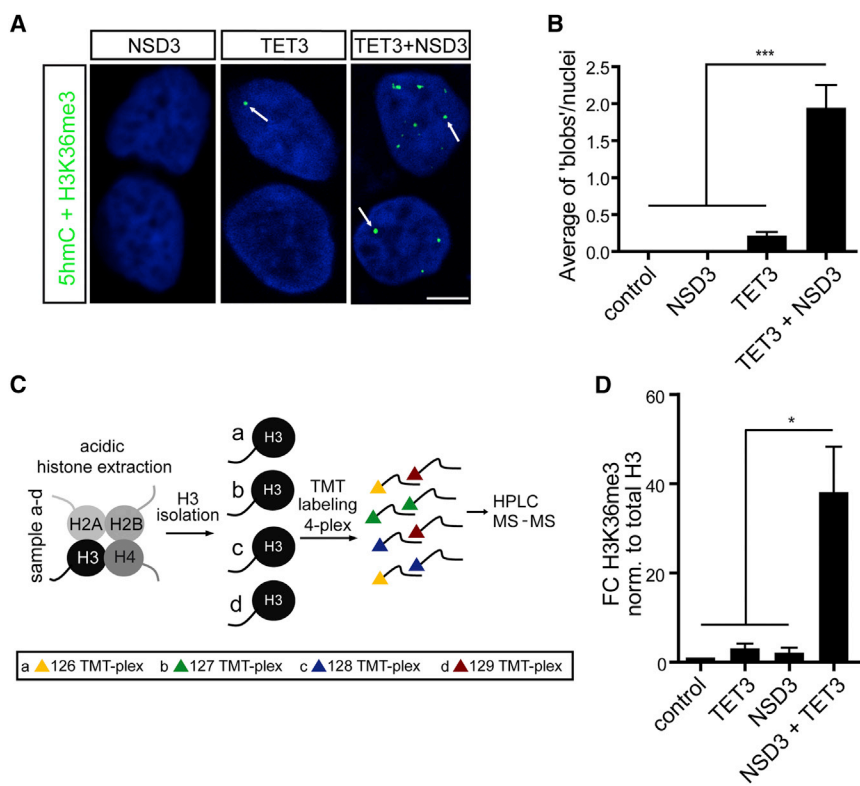


Figure 6. Effect of TET3-Mediated Hydroxylation on Chromatin Structure

(A) PLA signal for 5hmC and H3K36me3 in HEK293T cells overexpressing NSD3 (left), TET3-eGFP (middle) and TET3 + NSD3 (right). (B) Quantification data of the 5hmC/H3K36me3 PLA experiments in (A) are shown. (C) Scheme of TMT⁴-plex isobaric labeling and LC-MS quantification of H3 modifications is shown. (D) Isobaric labeling LC-MS quantification of H3K36me3 in HEK293T cells overexpressing TET3, NSD3, or TET3 and NSD3 compared to non-transfected cells. Summary data are mean \pm SEM in (B and D). * p < 0.05, *** p < 0.001 (one-way ANOVA). Scale bar, 5 μ m in (A).

tone H3 at lysine 36 (H3K36), which is an epigenetic mark commonly associated with transcription of active chromatin (Bernstein et al., 2005; Heintzman et al., 2007; Kouzarides, 2007).

TET3 Regulates NSD3 H3K36 Trimethylation Activity

To test if the TET3/NSD3 interaction is functionally relevant, we performed a PLA experiment with HEK293T cells transfected with TET3 and/or NSD3,

focusing on the co-localization of the products of the two enzymes, 5hmC and H3K36me3, respectively. NSD3 overexpression did not result in any detectable PLA signal (Figures 6A and 6B), which is in line with the very low levels of 5hmC in HEK293T cells (see Figure 4A). TET3-overexpressing cells showed only a few PLA signals, indicating that 5hmC elevation alone does not automatically result in 5hmC/H3K36me3 co-localization (Figures 6A and 6B). However, when both TET3 and NSD3 were overexpressed, we observed a synergism that resulted in a significant increase of the 5hmC/H3K36me3 PLA signal (Figures 6A and 6B). We next isolated and enriched histone 3 and quantified the extent of K36 trimethylation of histone 3 using isobaric labeling and MS (Figure 6C). This experiment showed a 40-fold increase in the levels of H3K36me3 in the HEK293T cells co-overexpressing TET3 and NSD3 compared to non-transfected cells (Figure 6D). Overexpression of TET3 or NSD3 alone provided only about a 2-fold increase of the H3K36me3 activation mark (Figure 6D). Thus, TET3 has a synergistic effect on the lysine methyltransferase activity of NSD3.

TET3 Overexpression in Retina Leads to Neuronal Gene Activation and H3K36me3

To analyze if TET3 overexpression has an effect on global gene expression in the retina, we transduced retinal explant cultures with lentiviral vectors expressing TET3-eGFP. Compared to retinal explant cultures overexpressing eGFP only, we identified 981 proteins that were upregulated after TET3-eGFP overexpression (Table S6). The TET3-eGFP overexpression strongly affects gene expression in the retina, resulting in the upregulation of proteins involved in several neurological functions (Figure 7A).

To test if REST has an effect on the genomic 5hmC content in the retina, we re-analyzed our 5hmC mapping data (see Figures 2A–2C) by searching for known REST target genes (Lu et al., 2014) within the subset of genes that gained 5hmC marks from week 2 to week 3 (Figure 5D). In support of a positive effect of REST on genomic 5hmC levels in the retina, we found a significant enrichment of REST target genes within the group of genes that gained 5hmC during retinal maturation ($p = 1.8 \times 10^{-7}$, Fischer's exact test) (Figure 5D).

Transcriptional activation of previously silenced genes requires chromatin remodeling, e.g., a switch from a heterochromatic to a euchromatic state (Voss and Hager, 2014). It was suggested that REST-mediated chromatin remodeling involves changes of the activating histone mark lysine 36 trimethylation of the nucleosomal histone H3 (H3K36me3) (Zheng et al., 2009). To test for this chromatin remodeling during retinal maturation, we analyzed the co-localization of 5hmC with H3K36me3 using specific antibodies in combination with PLA. As shown in Figures 5E and 5F, we observed a clear increase of the PLA signal for 5hmC/H3K36me3 co-localization in nuclei of 3-week-old retina, whereas at 2 weeks only a few or no PLA signals were observed. This increase in 5hmC/H3K36me3 PLA signal was observed in nuclei of all three major retinal layers (Figure 5F). Thus, from week 2 to week 3, the chromatin structure in retinal nuclei changed in a way that led to co-localization of 5hmC with the active chromatin mark H3K36me3.

Notably, three of the identified TET3-interacting proteins were the post-SET domain containing histone-lysine N-methyltransferases NSD3, NSD2, and SETD2 (Figure 3E). These histone writers are involved in the (tri-) methylation of nucleosomal his-

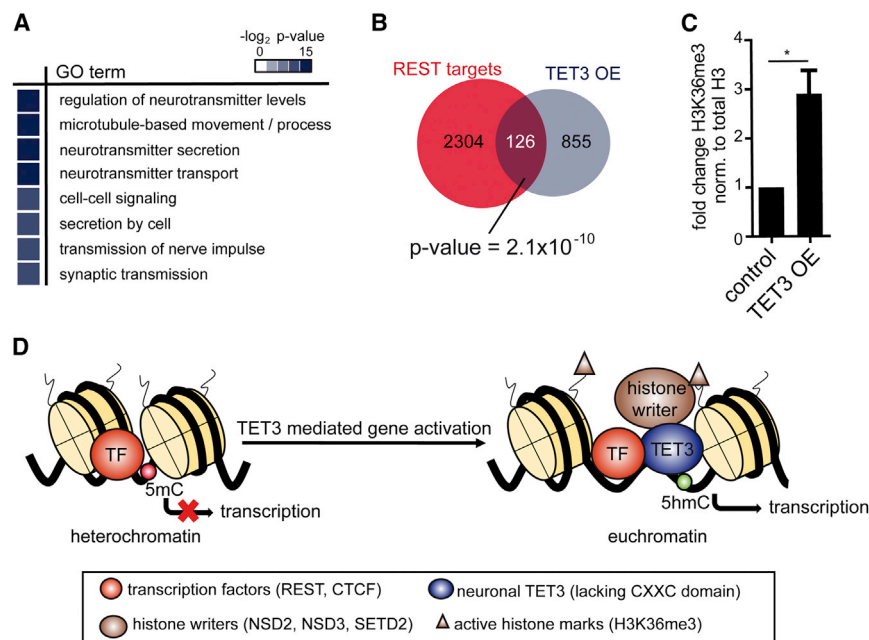


Figure 7. TET3 Overexpression Leads to Activation of Neuronal Genes and H3K36 Trimethylation during Retinal Maturation

(A) GO analysis reveals strong enrichment of neuronal proteins within the subset of proteins induced after TET3 OE in mouse retinal explants. (B) Venn diagram shows the significant overlap (Fisher's exact test; p value = 2.1×10^{-10}) of proteins enriched after TET3 OE and REST target genes.

(C) Isobaric labeling LC-MS quantification of H3K36me3 in mouse retina after TET3-eGFP or eGFP (control) OE. Summary data are mean \pm SEM. * $p < 0.05$ (Student's t test).

(D) Model for neuronal TET3-mediated transcriptional activation and chromatin remodeling. TET3 is recruited to the DNA by transcriptional regulators like REST for context-specific 5mC hydroxylation. Subsequently, TET3 mediates H3K36 trimethylation by recruitment of histone writers such as NSD3.

Correlation analysis again revealed that 126 known REST targets were significantly enriched in the subset of 981 TET3-upregulated proteins ($p = 2.1 \times 10^{-10}$, Fischer's exact test) (Figure 7B; Table S6).

To test whether TET3 also has a positive effect on lysine methyltransferase activity in the retina, we transduced retinal explant cultures with TET3-eGFP or eGFP using lentiviral vectors and determined the levels of H3K36me3 using isobaric labeling and MS. Overexpression of TET3-eGFP was able to elevate the levels of H3K36me3 by a factor of 3.8, suggesting that TET3-mediated hydroxylation of 5mC in retinal cells results in activation of NSD3 or similar enzymes to induce H3K36 trimethylation (Figure 7C). Our data indicate that this involves, at least in part, REST-assisted, TET3-mediated oxidation of genomic 5mC and subsequent NSD3-mediated chromatin remodeling.

DISCUSSION

Our comprehensive analysis of proteins interacting with the 5hmC-generating enzyme TET3 revealed a mechanism for specific epigenetic regulation of gene expression. We show that the transcription factor REST, a master regulator of genes involved in neuronal maturation processes (McGann et al., 2014), binds and recruits TET3 to the DNA for specific gene activation. In particular, we find that (1) REST regulates TET3 hydroxylase activity, (2) REST target genes accumulate 5hmC during retinal maturation, and (3) overexpression of TET3 activates REST target genes. In support of these findings, a recent study showed that deletion of REST in mouse embryonic stem cells (ESCs) led to significant loss of genome-wide 5hmC levels in regulatory regions (Feldmann et al., 2013). REST did not interact with TET1 or TET2. In addition, none of the TET3 peptides identified by MS after affinity purification of REST complexes from

mouse retinal lysates were specific for the TET3 isoforms with CXXC domain. Thus, REST-mediated recruitment of

TET3 seems to be specific for the neuronal TET3 isoform lacking a CXXC domain.

REST originally was described as a transcriptional repressor of neuronal genes in non-neuronal cells (Chong et al., 1995; Schoenherr and Anderson, 1995). In addition, REST was shown to be a key repressor in retinal ganglion cell commitment by preventing premature expression in retinal progenitor cells (Mao et al., 2011). It was hypothesized that, in neural progenitors, REST's repressive function is linked to gene regulatory networks related to neuronal differentiation (Mao et al., 2011). To act as a repressor, REST interacts with other proteins like mSin3a or COREST1 to form REST repressor complexes (Huang et al., 1999; McGann et al., 2014; Naruse et al., 1999). Importantly, none of the known REST-repressive complex proteins were identified in our REST-specific affinity purifications from mouse retinal lysates. There is now growing evidence suggesting that REST also plays a role in proper function of neuronal cells (Lu et al., 2014; Lunyak and Rosenfeld, 2005). For example, it was shown that REST functions as an activator of neuronal genes during neuronal maturation in the presence of a small double-stranded non-coding RNA corresponding to the NRSE/RE1 REST-binding sequence (Kuwabara et al., 2004).

Based on our 5hmC-mapping data, approximately 9% (521 of 5,921) of the genes that gained 5hmC during retinal maturation were known REST target genes. Thus, although REST contributes significantly ($p = 1.8 \times 10^{-7}$, Fischer's exact test) to context-dependent 5hmC generation, additional factors might exist for TET3 recruitment. In line with this, we identified several other transcriptional regulators interacting with TET3, such as ASXL1, CTCF, MORF4L1, SUZ12, and VAX1. It is tempting to speculate that these transcription regulators could influence TET3 hydroxylase activity in a similar way as REST. For instance, it was shown that the insulator CTCF interacts with TET enzymes

and that depletion of CTCF in mouse ESCs and adipocytes leads to significant loss of 5hmC in transcriptional enhancers (Dubois-Chevalier et al., 2014; Feldmann et al., 2013). However, CTCF seems not to discriminate among the TET isoforms, since it binds to TET1 and TET2 (Dubois-Chevalier et al., 2014) as well as to TET3 (present study).

Taken together, REST and most probably additional DNA-binding transcriptional regulators recruit TET3 to neuronal genes during retinal maturation to oxidize 5mC to 5hmC and activate their transcription. Transcriptional activation goes along with chromatin remodeling and translocation of genomic regions from 5mC-rich heterochromatin to 5hmC-rich euchromatin (Hahn et al., 2013; Mellén et al., 2012). Interestingly, our study reveals that TET3 interacts with several H3K36 methyltransferases. At a functional level TET3 stimulates the activity of NSD3 to generate the activating histone mark H3K36me3. Our data show that the transcriptional repressor REST can facilitate the recruitment of TET3 to 5mC-containing transcriptionally inactive genes (Figure 7D). This repressor-induced and context-specific recruitment is followed by TET3-mediated oxidation of 5mC to 5hmC. Subsequently, H3K36 methyltransferases induce chromatin remodeling to facilitate active transcription (Figure 7D).

Our findings favor a model that involves highly dynamic transcription-factor-dependent DNA hydroxymethylation that supports retinal network maturation by context-specific gene activation. Accordingly, our metabolic feeding experiments showed that the majority of nascent 5hmC marks were generated from pre-existing 5mC, supporting the idea of TET3-induced chromatin remodeling.

5hmC itself is known to recruit specific binding proteins (readers) (Mellén et al., 2012; Spruijt et al., 2013; Takai et al., 2014). We cannot exclude that such 5hmC readers might contribute to the observed TET3-mediated transcriptional activation during retinal maturation. Future study focusing on the identification of developmental-stage-specific 5hmC readers in the retina should help to elucidate their contribution.

In contrast to 5hmC, the levels of the further oxidized methylcytosine derivative 5fC in the retina were very low and decreased during neuronal maturation (see Figure S1C), suggesting that 5fC is a short-lived active DNA demethylation intermediate rather than an independent epigenetic mark in differentiating neuronal cells.

In conclusion, we provide evidence that TET3 and REST play a key role in the activation of neuronal gene transcription. Furthermore, we show that TET3-mediated gene activation involves H3K36 trimethylation-induced chromatin remodeling. This study provides a mechanistic link among REST-directed TET3-mediated 5hmC formation, subsequent chromatin remodeling by NSD3, and gene activation during neuronal maturation.

EXPERIMENTAL PROCEDURES

Animals

All mice used were on the C57BL/6 genetic background. All procedures concerning animals were performed with permission of the local authority (Regierung von Oberbayern). Day of birth was considered as postnatal day 1 (P1).

Lentiviral Constructs and Production

Mouse TET3 was PCR cloned from mouse retinal cDNA using the primers 5'-CTATCTAGAACCGCCATGGACTCAGGGCCAGTGAC-3' (forward) and 5'-TCACCGGTAAGATCCAGCGGCTGTAGGG-3' (reverse), and ligated into a lentiviral vector containing a cytomegalovirus (CMV) promoter (Mistrik et al., 2005), yielding LV-CMV-TET3ΔCXXC-eGFP. LV-CMV-eGFP expressing eGFP only was used for control experiments. Recombinant lentivirus as well as lentiviral particles were generated as described previously (Mistrik et al., 2005).

In Vitro Retinal Explant Cultures

Retinas from wild-type animals were used to generate retinal explants. Animals were killed and the eyes enucleated in DMEM Nutrient Mixture F12-HAM medium (Sigma) containing 10% fetal bovine serum (FBS) and antibiotics. Afterward, the entire eyes were incubated in DMEM-F12 serum-free medium containing 0.12% proteinase K at 37°C for 15 min, to allow preparation of retinal cultures with RPE attached. Subsequently, proteinase K was inactivated using 10% FBS in DMEM Nutrient Mixture F12-HAM medium. The eyes were dissected aseptically in a Petri dish. The anterior segment, lens, vitreous, sclera, and choroids were carefully removed, and the retina was cut perpendicular to its edges, resulting in a cloverleaf-like shape. Subsequently, the retina was transferred to a Millicell culture dish filter insert (Millipore) with the retinal pigment epithelium layer facing the membrane. The insert was put into a six-well culture plate and incubated in DMEM-F12 Nutrient medium at 37°C.

Transduction of Retinal Explant Cultures

Retinal explant cultures from 11-day-old mice were used for transduction experiments using LV-CMV-TET3ΔCXXC-eGFP (neuronal TET3) or LV-CMV-eGFP (control). Twelve retinæ were used for each experiment. The LV particles were added to the scleral and vitreal part of retinal explants directly after preparation of the explant cultures. Every second day, the full volume of DMEM-F12, 1.2 ml per dish, was replaced with fresh medium. Retinal explants were harvested after 12 days in vitro for nuclear protein extraction.

Cell Culture and Transfection

HEK293T cells were grown at 37°C and 10% CO₂ in DMEM supplemented with 10% FBS, 100 units/ml penicillin, and 100 µg/ml streptomycin. The cells were transfected with plasmids expressing mouse REST (Addgene plasmid 21310) and/or mouse TET3 (LV-CMV-TET3ΔCXXC-eGFP, see above) using calcium phosphate. Cells were incubated at 37°C and 10% CO₂ for 48 hr, with an additional medium exchange step after 24 hr from transfection. Upon harvesting, the cells were washed once with PBS after removing the medium, pelleted, and then lysed for subsequent genomic DNA extraction as described below.

For RNAi experiments, REST-specific siRNA duplexes or control siRNA duplexes (OriGene, SR304036) and plasmid-expressing mouse TET3 were transfected using siTRAN (OriGene) in Opti-MEM I reduced serum medium (Invitrogen) according to the manufacturer's instructions. Cells were incubated at 37°C and 10% CO₂ for 48 hr and subsequently harvested for genomic DNA isolation. Knockdown efficiency of gene expression was verified using qRT-PCR.

Tracing Experiments with [¹³C,³D₃]-Labeled L-methionine in Retinal Explants

De novo methylation-tracing experiments with [methyl-¹³C,³D₃]-methionine were performed in p11 retinal explant cultures (as described in the Supplemental Experimental Procedures) that were maintained in methionine-free DMEM containing 10% FBS and supplemented with 2 mM [methyl-¹³C,³D₃]-methionine for 9 days (after 3 days, media were renewed). Four retinal cultures were collected each 3 days after feeding and pooled for subsequent UHPLC-MS/MS quantification.

Data Interpretation of Interactors

From the identified proteins, only the proteins that were enriched in both the forward and the reverse experiments, with an enrichment factor of at least 1.5-fold over the control sample, are claimed as interactors. GO analysis

was performed using Ingenuity pathway analysis (IPA, Ingenuity Systems, <http://www.ingenuity.com/>) or DAVID bioinformatics database (<http://david.abcc.ncifcrf.gov/>).

Acid Extraction of Histones

To examine histone modifications of histone 3 (H3), we followed a protocol from Shechter et al. (2007). In brief, cells were lysed in 1 ml hypotonic buffer containing 10 mM Tris-HCl (pH 8.0), 1 mM KCl, 1.5 mM MgCl₂, and 1 mM DTT. Samples were incubated for 30 min on rotator at 4°C. The intact nuclei were centrifuged at 10,000 × *g* for 10 min at 4°C. Next, the nuclei were re-suspended in 400 μl 0.4 N H₂SO₄ and samples were incubated on rotator for 30 min. The nuclear debris was removed by centrifugation at 16,000 × *g* for 10 min. The supernatant was incubated for 30 min with 132 μl trichloroacetic acid (TCA) to precipitate histones. Histones were pelleted by 16,000 × *g* for 10 min and washed with acetone without disturbing the pellet to remove remaining acid from the solution. The pellets were eluted using ddH₂O.

Quantification of Histone Modifications

The visible protein band of histone H3 was cut out of a Coomassie-stained SDS-gel after destaining. The protein was in-gel alkylated and typically digested (Shevchenko et al., 2006). Trypsin activity was stopped using 1 mM PMSF and the peptides were labeled according to manufacturer's protocol with the TMT 2 (retina) or the TMT 4 (HEK293T cells).

Post-labeled samples of each experiment were pooled together and analyzed via LC-MS/MS. The relative reporter ion intensities of the trimethylated H3K36 peptide were normalized to the relative quantities of the whole protein. Finally, the fold change of the trimethylated peptides was calculated in relation to each control sample.

Bioinformatics

Sequence reads of 5hmC containing DNA fragments were mapped onto the reference mouse genome (NCBI Build UCSC mm9) using the Bowtie (v0.12.7) algorithm. Unique and monoclonal reads were used for further analysis. Refseq genes were downloaded from the UCSC mm9 annotation database (UCSC Genome Browser). GO analyses on differentially hydroxymethylated regions were performed using Ingenuity (IPA, Ingenuity Systems).

ACCESSION NUMBERS

The NCBI GEO accession number for the 5hmC profiling data reported in this paper is GSE65860. The ProteomeXchange accession number for the proteomics data reported in this paper is PXD001857.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, and 11 tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2015.03.020>.

AUTHOR CONTRIBUTIONS

S.M. and T.C. supervised the project. S.M., T.C., and A.P. designed experiments and interpreted data. A.P. conducted the analysis of gene expression, retinal explant culture, and LFQ proteomics. E.S., A.P., and N.M. performed protein pull-down studies. D.E. and S.K.L. performed peptide LC-MS quantification including sample preparation and analysis. A.F.K. conducted LFQ proteomics. M.W., J.S., and A.P. isolated genomic DNA. M.W., A.P., and J.S. performed sample preparation. M.W. and J.S. performed UHPLC-MS/MS quantification and isotope-tracing experiments of DNA bases. A.P., S.K., and S.M. performed and analyzed immunostainings and hMeDIP experiments. S.M. and E.S. generated expression plasmids and lentiviral vectors. A.P. and E.S. performed HEK293 cell culture experiments and acidic histone isolation. V.S. performed PLAs. A.P. and S.M. performed bioinformatics analysis. S.M. wrote the paper with input from other authors. M.B. and M.M. contributed to the experimental design, data interpretation, and manuscript editing.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (DFG) cluster of excellence (CIPSM, EXC114). A.F.K. is supported by a fellowship from the Peter und Traudl Engelhorn-Stiftung. We thank Gail Mandel for the gift of anti-REST antibodies; Tim M. Strom and Thomas Meitinger for NGS; Lukas Windhager, Ralf Zimmer, and Michael Bonin for help with bioinformatics analysis; and Fred Koch for help with cloning and qPCR.

Received: November 7, 2014

Revised: January 9, 2015

Accepted: March 7, 2015

Published: April 2, 2015

REFERENCES

- Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 120, 169–181.
- Chong, J.A., Tapia-Ramirez, J., Kim, S., Toledo-Aral, J.J., Zheng, Y., Boutros, M.C., Altschuler, Y.M., Frohman, M.A., Kraner, S.D., and Mandel, G. (1995). REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* 80, 949–957.
- Colquitt, B.M., Allen, W.E., Barnea, G., and Lomvardas, S. (2013). Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cell identity. *Proc. Natl. Acad. Sci. USA* 110, 14682–14687.
- Dietrich, N., Lerdrup, M., Landt, E., Agrawal-Singh, S., Bak, M., Tommerup, N., Rappsilber, J., Södersten, E., and Hansen, K. (2012). REST-mediated recruitment of polycomb repressor complexes in mammalian cells. *PLoS Genet.* 8, e1002494.
- Dubois-Chevalier, J., Oger, F., Dehondt, H., Firmin, F.F., Gheeraert, C., Staels, B., Lefebvre, P., and Eeckhoutte, J. (2014). A dynamic CTCF chromatin binding landscape promotes DNA hydroxymethylation and transcriptional induction of adipocyte differentiation. *Nucleic Acids Res.* 42, 10943–10959.
- Feldmann, A., Ivanek, R., Murr, R., Gaidatzis, D., Burger, L., and Schübeler, D. (2013). Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet.* 9, e1003994.
- Gregory-Evans, C.Y., Wallace, V.A., and Gregory-Evans, K. (2013). Gene networks: dissecting pathways in retinal development and disease. *Prog. Retin. Eye Res.* 33, 40–66.
- Hahn, M.A., Qiu, R., Wu, X., Li, A.X., Zhang, H., Wang, J., Jui, J., Jin, S.G., Jiang, Y., Pfeifer, G.P., and Lu, Q. (2013). Dynamics of 5-hydroxymethylcytosine and chromatin marks in Mammalian neurogenesis. *Cell Rep.* 3, 291–300.
- Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39, 311–318.
- Hoon, M., Okawa, H., Della Santina, L., and Wong, R.O. (2014). Functional architecture of the retina: development and disease. *Prog. Retin. Eye Res.* 42, 44–84.
- Huang, Y., Myers, S.J., and Dingleline, R. (1999). Transcriptional repression by REST: recruitment of Sin3A and histone deacetylase to neuronal genes. *Nat. Neurosci.* 2, 867–872.
- Ko, M., An, J., Bandukwala, H.S., Chavez, L., Aijö, T., Pastor, W.A., Segal, M.F., Li, H., Koh, K.P., Lähdesmäki, H., et al. (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* 497, 122–126.
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell* 128, 693–705.
- Kraner, S.D., Chong, J.A., Tsay, H.J., and Mandel, G. (1992). Silencing the type II sodium channel gene: a model for neural-specific gene regulation. *Neuron* 9, 37–44.

- Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930.
- Kuwabara, T., Hsieh, J., Nakashima, K., Taira, K., and Gage, F.H. (2004). A small modulatory dsRNA specifies the fate of adult neural stem cells. *Cell* 116, 779–793.
- Liu, N., Wang, M., Deng, W., Schmidt, C.S., Qin, W., Leonhardt, H., and Spada, F. (2013). Intrinsic and extrinsic connections of Tet3 dioxygenase with CXXC zinc finger modules. *PLoS ONE* 8, e62755.
- Lu, T., Aron, L., Zullo, J., Pan, Y., Kim, H., Chen, Y., Yang, T.H., Kim, H.M., Drake, D., Liu, X.S., et al. (2014). REST and stress resistance in ageing and Alzheimer's disease. *Nature* 507, 448–454.
- Lunyak, V.V., and Rosenfeld, M.G. (2005). No rest for REST: REST/NRSF regulation of neurogenesis. *Cell* 121, 499–501.
- Mao, C.A., Tsai, W.W., Cho, J.H., Pan, P., Barton, M.C., and Klein, W.H. (2011). Neuronal transcriptional repressor REST suppresses an Atoh7-independent program for initiating retinal ganglion cell development. *Dev. Biol.* 349, 90–99.
- McGann, J.C., Oyer, J.A., Garg, S., Yao, H., Liu, J., Feng, X., Liao, L., Yates, J.R., 3rd, and Mandel, G. (2014). Polycarbonyl- and REST-associated histone deacetylases are independent pathways toward a mature neuronal phenotype. *eLife* 3, e04235.
- Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 151, 1417–1430.
- Mistrić, P., Mader, R., Michalakakis, S., Weidinger, M., Pfeifer, A., and Biel, M. (2005). The murine HCN3 gene encodes a hyperpolarization-activated cation channel with slow kinetics and unique response to cyclic nucleotides. *J. Biol. Chem.* 280, 27056–27061.
- Münzel, M., Globisch, D., Brückl, T., Wagner, M., Welzmler, V., Michalakakis, S., Müller, M., Biel, M., and Carell, T. (2010). Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.* 49, 5375–5377.
- Naruse, Y., Aoki, T., Kojima, T., and Mori, N. (1999). Neural restrictive silencer factor recruits mSin3 and histone deacetylase complex to repress neuron-specific target genes. *Proc. Natl. Acad. Sci. USA* 96, 13691–13696.
- Okawa, H., Hoon, M., Yoshimatsu, T., Della Santina, L., and Wong, R.O. (2014). Illuminating the multifaceted roles of neurotransmission in shaping neuronal circuitry. *Neuron* 83, 1303–1318.
- Santiago, M., Antunes, C., Guedes, M., Sousa, N., and Marques, C.J. (2014). TET enzymes and DNA hydroxymethylation in neural development and function - how critical are they? *Genomics* 104, 334–340.
- Schiesser, S., Pfaffeneder, T., Sadeghian, K., Hackner, B., Steigenberger, B., Schröder, A.S., Steinbacher, J., Kashiwazaki, G., Höfner, G., Wanner, K.T., et al. (2013). Deamination, oxidation, and C-C bond cleavage reactivity of 5-hydroxymethylcytosine, 5-formylcytosine, and 5-carboxycytosine. *J. Am. Chem. Soc.* 135, 14593–14599.
- Schoenherr, C.J., and Anderson, D.J. (1995). The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* 267, 1360–1363.
- Shechter, D., Dormann, H.L., Allis, C.D., and Hake, S.B. (2007). Extraction, purification and analysis of histones. *Nat. Protoc.* 2, 1445–1457.
- Shevchenko, A., Tomas, H., Havlis, J., Olsen, J.V., and Mann, M. (2006). In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* 1, 2856–2860.
- Söderberg, O., Gullberg, M., Jarvius, M., Ridderstråle, K., Leuchowius, K.J., Jarvius, J., Wester, K., Hydbring, P., Bahram, F., Larsson, L.G., and Landegren, U. (2006). Direct observation of individual endogenous protein complexes in situ by proximity ligation. *Nat. Methods* 3, 995–1000.
- Solovei, I., Kreysing, M., Lancôt, C., Kösem, S., Peichl, L., Cremer, T., Guck, J., and Joffe, B. (2009). Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution. *Cell* 137, 356–368.
- Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., Zhang, W., Jian, X., et al. (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* 29, 68–72.
- Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., et al. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152, 1146–1159.
- Szulwach, K.E., Li, X., Li, Y., Song, C.X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M., Levey, A.I., et al. (2011). 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* 14, 1607–1616.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935.
- Takai, H., Masuda, K., Sato, T., Sakaguchi, Y., Suzuki, T., Suzuki, T., Koyama-Nasu, R., Nasu-Nishimura, Y., Katou, Y., Ogawa, H., et al. (2014). 5-Hydroxymethylcytosine plays a critical role in glioblastomagenesis by recruiting the CHTOP-methylosome complex. *Cell Rep.* 9, 48–60.
- Voss, T.C., and Hager, G.L. (2014). Dynamic regulation of transcriptional states by chromatin and transcription factors. *Nat. Rev. Genet.* 15, 69–81.
- Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A., Rappaport, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 473, 343–348.
- Xiang, M. (2013). Intrinsic control of mammalian retinogenesis. *Cell. Mol. Life Sci.* 70, 2519–2532.
- Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S., et al. (2011). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell* 42, 451–464.
- Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J.G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T., et al. (2012). Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for *Xenopus* eye and neural development. *Cell* 151, 1200–1213.
- Zheng, D., Zhao, K., and Mehler, M.F. (2009). Profiling RE1/REST-mediated histone modifications in the human genome. *Genome Biol.* 10, R9.

5 Abkürzungsverzeichnis

A	Adenin oder Adenosin
AID	<i>Activation-induced deaminase</i>
Am	2'-O-Methyladenosin
Ar	2'-O-Ribosyladenosin
ASL	<i>Anticodon stem and loop</i>
BER	Basenexzisionsreparatur
C	Cytosin oder Cytidin
caC	5-Carboxycytosin
Cm	2'-O-Methylcytosin
CpG	dC-dG Dinukleotid
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Desoxyribonukleinsäure
ds	doppelsträngig
DSL	<i>Dihydrouridine stem and loop</i>
<i>E.coli</i>	<i>Escherichia coli</i>
ESCs	<i>Embryonic stem cells</i>
ESI	Elektrospray-Ionisation
fC	5-Formylcytosin
Gm	2'-O-Methylguanosin
h	Stunde(n)
hmC	5-Hydroxymethylcytosin
hmU	5-Hydroxymethyluracil
HPLC	<i>High performance liquid chromatography</i>
i ⁶ A	N ⁶ -Isopentenyladenosin
m ¹ A	1-Methyladenosin
m ¹ G	1-Methylguanosin
m ² A	2-Methyladenosin
m ² G	N ² -Methylguanosin
m ² ₂ G	N ² ,N ² -Dimethylguanosine
m ⁷ G	7-Methylguanosin
mC	5-Methylcytosin
min	Minute(n)
MS	Massenspektrometrie
ms ² i ⁶ A	2-Methylthio-N ⁶ -isopentenyladenosin
OHyW	Hydroxywybutosin
PCR	<i>Polymerase chain reaction</i>
Q	Queuosin
RNA	Ribonukleinsäure
<i>S. cerevisiae</i>	<i>Saccharomyces cerevisiae</i>
SAM	S-Adenosylmethionin
SMUG1	Single-strand-selective monofunctional uracil-DNA glycosylase 1

T	Thymin oder Thymidin
TDG	Thymin-DNA-Glycosylase
t ⁶ A	N ⁶ -Theronylcarbamoyladenosen
Tet	Ten-eleven translocation
tRNA	Transfer-Ribonukleinsäure
TSL	<i>Thymidine stem and loop</i>
U	Uracil oder Uridin
ψ	Pseudouridin
yW	Wybutosin

6 Literaturverzeichnis

- [1] Frankfurter Allgemeine Zeitung vom 27. Juni 2000.
- [2] T. Carell, C. Brandmayr, A. Hienzsch, M. Müller, D. Pearson, V. Reiter, I. Thoma, P. Thumbs, M. Wagner, *Angew. Chem. Int. Ed.* **2012**, *51*, 7110-7131.
- [3] Y. Motorin, M. Helm, *Wiley Interdiscip. Rev. RNA* **2011**, *2*, 611-631.
- [4] A. Hienzsch, C. Deiml, V. Reiter, T. Carell, *Chem. Eur. J.* **2013**, *19*, 4244-4248.
- [5] P. F. Agris, *Nucleic Acids Res.* **2004**, *32*, 223-238.
- [6] W. A. Cantara, P. F. Crain, J. Rozenski, J. A. McCloskey, K. A. Harris, X. Zhang, F. A. P. Vendeix, D. Fabris, P. F. Agris, *Nucleic Acids Res.* **2011**, *39*, D195-D201.
- [7] M. A. Machnicka, K. Milanowska, O. O. Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Horkawicz, K. M. Rother, M. Helm, J. M. Bujnicki, H. Grosjean, *Nucleic Acids Res.* **2013**, *41*, D262-D267.
- [8] B. El Yacoubi, M. Bailly, V. de Crecy-Lagard, *Annu. Rev. Genet.* **2012**, *46*, 69-95.
- [9] F. H. Crick, L. Barnett, S. Brenner, R. J. Watts-Tobin, *Nature* **1961**, *192*, 1227-1232.
- [10] E. M. Novoa, L. Ribas de Pouplana, *Trends Genet.* **2012**, *28*, 574-581.
- [11] P. F. Agris, *EMBO Rep.* **2008**, *9*, 629-635.
- [12] M. Helm, *Nucleic Acids Res.* **2006**, *34*, 721-733.
- [13] Y. Motorin, M. Helm, *Biochemistry* **2010**, *49*, 4934-4944.
- [14] P. J. Beuning, K. Musier-Forsyth, *Biopolymers* **1999**, *52*, 1-28.
- [15] G. R. Wyatt, *Nature* **1950**, *166*, 237-238.
- [16] R. J. Klose, A. P. Bird, *Trends Biochem. Sci.* **2006**, *31*, 89-97.
- [17] E. Li, Y. Zhang, *Cold Spring Harb. Perspect. Biol.* **2014**, *6*:a019133.
- [18] D. Globisch, M. Münzel, M. Müller, S. Michalakis, M. Wagner, S. Koch, T. Brückl, M. Biel, T. Carell, *PLoS ONE* **2010**, *5*:e15367.
- [19] A. M. Deaton, A. Bird, *Genes Dev.* **2011**, *25*, 1010-1022.
- [20] P. A. Jones, *Nat. Rev. Genet.* **2012**, *13*, 484-492.
- [21] S. Saxonov, P. Berg, D. L. Brutlag, *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 1412-1417.
- [22] S. Kriaucionis, N. Heintz, *Science* **2009**, *324*, 929-930.
- [23] M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, A. Rao, *Science* **2009**, *324*, 930-935.
- [24] N. W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski, R. Yura, *Biochem. J.* **1972**, *126*, 781-790.
- [25] T. Pfaffeneder, B. Hackner, M. Truss, M. Münzel, M. Müller, C. A. Deiml, C. Hagemeier, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 7008-7012.
- [26] S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, Y. Zhang, *Science* **2011**, *333*, 1300-1303.
- [27] Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He, G. L. Xu, *Science* **2011**, *333*, 1303-1307.
- [28] H. Wu, Y. Zhang, *Genes Dev.* **2011**, *25*, 2436-2452.
- [29] W. A. Pastor, L. Aravind, A. Rao, *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 341-356.
- [30] R. M. Kohli, Y. Zhang, *Nature* **2013**, *502*, 472-479.
- [31] T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schürmann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Müller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schär, M. Müller, T. Carell, *Nat. Chem. Biol.* **2014**, *10*, 574-581.
- [32] J. F. Mouret, M. Polverelli, F. Sarrazini, J. Cadet, *Chem. Biol. Interact.* **1991**, *77*, 187-201.
- [33] J. U. Guo, Y. Su, C. Zhong, G. L. Ming, H. Song, *Cell* **2011**, *145*, 423-434.
- [34] A. Maiti, A. C. Drohat, *J. Biol. Chem.* **2011**, *286*, 35334-35338.
- [35] S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue, A. Bellacosa, *Cell* **2011**, *146*, 67-79.
- [36] S. C. Wu, Y. Zhang, *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 607-620.
- [37] D. M. Messerschmidt, B. B. Knowles, D. Solter, *Genes Dev.* **2014**, *28*, 812-828.
- [38] S. Kar, S. Parbin, M. Deb, A. Shilpi, D. Sengupta, S. K. Rath, M. Rakshit, A. Patra, S. K. Patra, *Cell. Mol. Life Sci.* **2014**, *71*, 1017-1032.

-
- [39] X. Li, W. Wei, V. S. Ratnu, T. W. Bredy, *Neurobiol. Learn. Mem.* **2013**, *105*, 125-132.
- [40] M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 1049-1055.
- [41] C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. T. C. Jansen, C. Bauer, M. Münzel, M. Wagner, M. Müller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Müller, J. Walter, R. Böhlens, H. van Ingen, H. Leonhardt, T. Carell, M. Vermeulen, *Cell* **2013**, *152*, 1146-1159.
- [42] J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, et al., *Science* **2001**, *291*, 1304-1351.
- [43] E. S. Lander, L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczy, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chisoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al., *Nature* **2001**, *409*, 860-921.
- [44] A. D. Goldberg, C. D. Allis, E. Bernstein, *Cell* **2007**, *128*, 635-638.
- [45] S. Feng, S. E. Jacobsen, W. Reik, *Science* **2010**, *330*, 622-627.
- [46] I. Cantone, A. G. Fisher, *Nat. Struct. Mol. Biol.* **2013**, *20*, 282-289.
- [47] S. Kriaucionis, M. Tahiliani, *Cold Spring Harb. Perspect. Biol.* **2014**, *6*:a018630.
- [48] N. Liu, T. Pan, *Transl. Res.* **2015**, *165*, 28-35.
- [49] H. Wu, Y. Zhang, *Cell* **2014**, *156*, 45-68.
- [50] C. Jiang, B. F. Pugh, *Nat. Rev. Genet.* **2009**, *10*, 161-172.
- [51] A. R. Cutter, J. J. Hayes, *FEBS Lett.* **2015**, DOI:10.1016/j.febslet.2015.05.016.
- [52] J. Wu, M. Grunstein, *Trends Biochem. Sci.* **2000**, *25*, 619-623.
- [53] A. F. Kebede, R. Schneider, S. Daujat, *FEBS J.* **2015**, *282*, 1658-1674.
- [54] S. B. Rothbart, B. D. Strahl, *Biochim. Biophys. Acta* **2014**, *1839*, 627-643.
- [55] P. Tessarz, T. Kouzarides, *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 703-708.
- [56] X. Cheng, *Cold Spring Harb. Perspect. Biol.* **2014**, DOI: 10.1101/cshperspect.a018747.
- [57] K. S. Leung, V. W. Cheng, S. W. Mok, S. K. Tsui, *Curr. Stem Cell Res. Ther.* **2014**, *9*, 388-395.
- [58] N. R. Rose, R. J. Klose, *Biochim. Biophys. Acta* **2014**, *1839*, 1362-1372.
- [59] G. F. Jia, Y. Fu, X. Zhao, Q. Dai, G. Q. Zheng, Y. Yang, C. Q. Yi, T. Lindahl, T. Pan, Y. G. Yang, C. He, *Nat. Chem. Biol.* **2011**, *7*, 885-887.
- [60] Y. Yue, J. Liu, C. He, *Genes Dev.* **2015**, *29*, 1343-1355.
- [61] X. Wang, B. S. Zhao, I. A. Roundtree, Z. Lu, D. Han, H. Ma, X. Weng, K. Chen, H. Shi, C. He, *Cell* **2015**, *161*, 1388-1399.
- [62] C. A. He, *Nat. Chem. Biol.* **2010**, *6*, 863-865.
- [63] T. Pan, *Trends Biochem. Sci.* **2013**, *38*, 204-209.
- [64] M. Lee, B. Kim, V. N. Kim, *Cell* **2014**, *158*, 980-987.
- [65] K. D. Meyer, S. R. Jaffrey, *Nat. Rev. Mol. Cell Biol.* **2014**, *15*, 313-326.
- [66] U. Chandola, R. Das, B. Panda, *Brief. Funct. Genomics.* **2015**, *14*, 169-179.
- [67] S. Horowitz, A. Horowitz, T. W. Nilsen, T. W. Munns, F. M. Rottman, *P. Natl. Acad. Sci.-Biol.* **1984**, *81*, 5667-5671.

-
- [68] T. Sibbritt, H. R. Patel, T. Preiss, *Wiley Interdiscip. Rev. RNA* **2013**, *4*, 397-422.
 - [69] B. Linder, A. V. Grozhik, A. O. Olarerin-George, C. Meydan, C. E. Mason, S. R. Jaffrey, *Nat. Methods* **2015**, *12*, 767-772.
 - [70] J. A. Bokar, M. E. Shambaugh, D. Polayes, A. G. Matera, F. M. Rottman, *RNA* **1997**, *3*, 1233-1247.
 - [71] J. Z. Liu, Y. N. Yue, D. L. Han, X. Wang, Y. Fu, L. Zhang, G. F. Jia, M. Yu, Z. K. Lu, X. Deng, Q. Dai, W. Z. Chen, C. He, *Nat. Chem. Biol.* **2014**, *10*, 93-95.
 - [72] X. L. Ping, B. F. Sun, L. Wang, W. Xiao, X. Yang, W. J. Wang, S. Adhikari, Y. Shi, Y. Lv, Y. S. Chen, X. Zhao, A. Li, Y. Yang, U. Dahal, X. M. Lou, X. Liu, J. Huang, W. P. Yuan, X. F. Zhu, T. Cheng, Y. L. Zhao, X. Q. Wang, J. M. R. Danielsen, F. Liu, Y. G. Yang, *Cell Res.* **2014**, *24*, 177-189.
 - [73] G. Zheng, J. A. Dahl, Y. Niu, P. Fedorcsak, C. M. Huang, C. J. Li, C. B. Vagbo, Y. Shi, W. L. Wang, S. H. Song, Z. Lu, R. P. Bosmans, Q. Dai, Y. J. Hao, X. Yang, W. M. Zhao, W. M. Tong, X. J. Wang, F. Bogdan, K. Furu, Y. Fu, G. Jia, X. Zhao, J. Liu, H. E. Krokan, A. Klungland, Y. G. Yang, C. He, *Mol. Cell* **2013**, *49*, 18-29.
 - [74] G. Zheng, J. A. Dahl, Y. Niu, Y. Fu, A. Klungland, Y. G. Yang, C. He, *RNA Biol.* **2013**, *10*, 915-918.
 - [75] T. Gerken, C. A. Girard, Y. C. L. Tung, C. J. Webby, V. Saudek, K. S. Hewitson, G. S. H. Yeo, M. A. McDonough, S. Cunliffe, L. A. McNeill, J. Galvanovskis, P. Rorsman, P. Robins, X. Prieur, A. P. Coll, M. Ma, Z. Jovanovic, I. S. Farooqi, B. Sedgwick, I. Barroso, T. Lindahl, C. P. Ponting, F. M. Ashcroft, S. O'Rahilly, C. J. Schofield, *Science* **2007**, *318*, 1469-1472.
 - [76] W. Chen, L. Zhang, G. Zheng, Y. Fu, Q. Ji, F. Liu, H. Chen, C. He, *FEBS Lett.* **2014**, *588*, 892-898.
 - [77] A. Thalhammer, Z. Bencokova, R. Poole, C. Loenarz, J. Adam, L. O'Flaherty, J. Schodel, D. Mole, K. Giaslakitiotis, C. J. Schofield, E. M. Hammond, P. J. Ratcliffe, P. J. Pollard, *PLoS ONE* **2011**, *6*, DOI:10.1371/journal.pone.0016210.
 - [78] B. Delatte, R. Deplus, F. Fuks, *EMBO J.* **2014**, *33*, 1198-1211.
 - [79] Y. Fu, G. F. Jia, X. Q. Pang, R. N. Wang, X. Wang, C. J. Li, S. Smemo, Q. Dai, K. A. Bailey, M. A. Nobrega, K. L. Han, Q. Cui, C. He, *Nat. Commun.* **2013**, *4*, DOI:10.1038/ncomms2822.
 - [80] X. Wang, C. He, *RNA Biol.* **2014**, *11*, 669-672.
 - [81] D. Dominissini, S. Moshitch-Moshkovitz, S. Schwartz, M. Salmon-Divon, L. Ungar, S. Osenberg, K. Cesarkas, J. Jacob-Hirsch, N. Amariglio, M. Kupiec, R. Sorek, G. Rechavi, *Nature* **2012**, *485*, 201-206.
 - [82] W. J. Ma, S. Cheng, C. Campbell, A. Wright, H. Furneaux, *J. Biol. Chem.* **1996**, *271*, 8144-8151.
 - [83] P. Kundu, M. R. Fabian, N. Sonenberg, S. N. Bhattacharyya, W. Filipowicz, *Nucleic Acids Res.* **2012**, *40*, 5088-5100.
 - [84] Y. Wang, Y. Li, J. I. Toth, M. D. Petroski, Z. L. Zhang, J. C. Zhao, *Nat. Cell Biol.* **2014**, *16*, 191-198.
 - [85] P. Stoilov, I. Rafalska, S. Stamm, *Trends Biochem. Sci.* **2002**, *27*, 495-497.
 - [86] X. Wang, Z. Lu, A. Gomez, G. C. Hon, Y. Yue, D. Han, Y. Fu, M. Parisien, Q. Dai, G. Jia, B. Ren, T. Pan, C. He, *Nature* **2014**, *505*, 117-120.
 - [87] G. T. a. E. M. Mutation Research, C., S. Lee, E. A. Bagg, J. S. McTaggart, R. Deacon, T. Gerken, A. Lee, L. Moir, J. Mecinovic, M. M. Quwailid, C. J. Schofield, F. M. Ashcroft, R. D. Cox, *PLoS Genet.* **2009**, *5*, e1000599.
 - [88] T. M. Frayling, N. J. Timpson, M. N. Weedon, E. Zeggini, R. M. Freathy, C. M. Lindgren, J. R. B. Perry, K. S. Elliott, H. Lango, N. W. Rayner, B. Shields, L. W. Harries, J. C. Barrett, S. Ellard, C. J. Groves, B. Knight, A. M. Patch, A. R. Ness, S. Ebrahim, D. A. Lawlor, S. M. Ring, Y. Ben-Shlomo, M. R. Jarvelin, U. Sovio, A. J. Bennett, D. Melzer, L. Ferrucci, R. J. F. Loos, I. Barroso, N. J. Wareham, F. Karpe, K. R. Owen, L. R. Cardon, M. Walker, G. A. Hitman, C. N. A. Palmer, A. S. F. Doney, A. D. Morris, G. D. Smith, A. T. Hattersley, M. I. McCarthy, C. Wellcome Trust Case, *Science* **2007**, *316*, 889-894.
 - [89] E. Karra, O. G. O'Daly, A. I. Choudhury, A. Yousseif, S. Millership, M. T. Neary, W. R. Scott, K. Chandarana, S. Manning, M. E. Hess, H. Iwakura, T. Akamizu, Q. Millet, C. Gelegen, M. E. Drew, S. Rahman, J. J. Emmanuel, S. C. R. Williams, U. U. Ruther, J. C. Bruning, D. J. Withers, F. O. Zelaya, R. L. Batterham, *J. Clin. Investig.* **2013**, *123*, 3539-3551.
 - [90] E. Zeggini, M. N. Weedon, C. M. Lindgren, T. M. Frayling, K. S. Elliott, H. Lango, N. J. Timpson, J. R. B. Perry, N. W. Rayner, R. M. Freathy, J. C. Barrett, B. Shields, A. P. Morris, S. Ellard, C. J. Groves, L. W. Harries, J. L. Marchini, K. R. Owen, B. Knight, L. R. Cardon, M. Walker, G. A. Hitman, A. D. Morris, A. S. F. Doney, M. I. McCarthy, A. T. Hattersley, *Science* **2007**, *316*, 1336-1341.
 - [91] I. Kalnina, L. Zaharenko, I. Vaivade, V. Rovite, L. Nikitina-Zake, R. Peculis, D. Fridmanis, K. Geldnere, J. A. Jacobsson, M. S. Almen, V. Pirags, H. B. Schioth, J. Klovins, *Gene* **2013**, *527*, 462-468.
 - [92] T. M. Carlile, M. F. Rojas-Duran, B. Zinshteyn, H. Shin, K. M. Bartoli, W. V. Gilbert, *Nature* **2014**, *515*, 143-146.
 - [93] S. Schwartz, D. A. Bernstein, M. R. Mumbach, M. Jovanovic, R. H. Herbst, B. X. Leon-Ricardo, J. M. Engreitz, M. Guttman, R. Satija, E. S. Lander, G. Fink, A. Regev, *Cell* **2014**, *159*, 148-162.
 - [94] X. Li, P. Zhu, S. Ma, J. Song, J. Bai, F. Sun, C. Yi, *Nat. Chem. Biol.* **2015**, *11*, 592-597.

-
- [95] F. Spenkuch, Y. Motorin, M. Helm, *RNA Biol.* **2014**, *11*, 1540-1554.
- [96] B. S. Zhao, C. He, *Cell Res.* **2015**, *25*, 153-154.
- [97] T. Brückl, D. Globisch, M. Wagner, M. Müller, T. Carell, *Angew. Chem. Int. Ed.* **2009**, *48*, 7932-7934.
- [98] T. Brückl, PhD thesis, Ludwig-Maximilians-Universität München (Munich, Germany), **2010**.
- [99] A. Hienzs, PhD thesis, Ludwig-Maximilians-Universität München (Munich, Germany), **2012**.
- [100] C. Brandmayr, M. Wagner, T. Brueckl, D. Globisch, D. Pearson, A. C. Kneutinger, V. Reiter, A. Hienzs, S. Koch, I. Thoma, P. Thumbs, S. Michalakis, M. Mueller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 11162-11165.
- [101] L. M. Iyer, M. Tahiliani, A. Rao, L. Aravind, *Cell Cycle* **2009**, *8*, 1698-1710.
- [102] X. Lu, B. S. Zhao, C. He, *Chem. Rev.* **2015**, *115*, 2225-2239.
- [103] R. M. Kohli, Y. Zhang, *Nature* **2013**, *502*, 472-479.
- [104] M. Okano, D. W. Bell, D. A. Haber, E. Li, *Cell* **1999**, *99*, 247-257.
- [105] J. Song, M. Teplova, S. Ishibe-Murakami, D. J. Patel, *Science* **2012**, *335*, 709-712.
- [106] I. Cantone, A. G. Fisher, *Nat. Struct. Mol. Biol.* **2013**, *20*, 282-289.
- [107] L. Tan, Y. G. Shi, *Development* **2012**, *139*, 1895-1902.
- [108] H. Zhao, T. Chen, *J. Hum. Genet.* **2013**, *58*, 421-427.
- [109] S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, Y. Zhang, *Nature* **2010**, *466*, 1129-1133.
- [110] J. Cadet, J. R. Wagner, *Mutat. Res.-Gen. Tox. En.* **2014**, *764-765*, 18-35.
- [111] B. Delatte, R. Deplus, F. Fuks, *EMBO J* **2014**, *33*, 1198-1211.
- [112] F. P. Güngerich, *J. Biol. Chem.* **2015**, *290*, 20700-20701.
- [113] M. Münzel, D. Globisch, T. Brückl, M. Wagner, V. Welzmler, S. Michalakis, M. Müller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2010**, *49*, 5375-5377.
- [114] T. F. J. Kraus, D. Globisch, M. Wagner, S. Eigenbrod, D. Widmann, M. Münzel, M. Müller, T. Pfaffeneder, B. Hackner, W. Feiden, U. Schüller, T. Carell, H. A. Kretzschmar, *Int. J. Cancer* **2012**, *131*, 1577-1590.
- [115] M. Wagner, J. Steinbacher, T. F. Kraus, S. Michalakis, B. Hackner, T. Pfaffeneder, A. Perera, M. Müller, A. Giese, H. A. Kretzschmar, T. Carell, *Angew. Chem. Int. Ed.* **2015**, DOI: 10.1002/anie.201502722.
- [116] M. Dizdaroğlu, D. S. Bergtold, *Anal. Biochem.* **1986**, *156*, 182-188.
- [117] D. W. Potter, Z. Djuric, *Adv. Exp. Med. Biol.* **1991**, *283*, 801-803.
- [118] L. M. Iyer, S. Abhiman, L. Aravind, *Prog. Mol. Biol. Transl. Sci.* **2011**, *101*, 25-104.
- [119] N. Liu, M. Wang, W. Deng, C. S. Schmidt, W. Qin, H. Leonhardt, F. Spada, *PLoS ONE* **2013**, *8*, e62755.
- [120] L. Hu, Z. Li, J. Cheng, Q. Rao, W. Gong, M. Liu, Y. G. Shi, J. Zhu, P. Wang, Y. Xu, *Cell* **2013**, *155*, 1545-1555.
- [121] H. Hashimoto, J. E. Pais, X. Zhang, L. Saleh, Z. Q. Fu, N. Dai, I. R. Correa, Jr., Y. Zheng, X. Cheng, *Nature* **2014**, *506*, 391-395.
- [122] H. Hashimoto, J. E. Pais, N. Dai, I. R. Correa, Jr., X. Zhang, Y. Zheng, X. Cheng, *Nucleic Acids Res.* **2015**, doi:10.1093/nar/gkv870.
- [123] H. Hashimoto, X. Zhang, P. M. Vertino, X. Cheng, *J. Biol. Chem.* **2015**, *290*, 20723-20733.
- [124] H. K. Long, N. P. Blackledge, R. J. Klose, *Biochem. Soc. Trans.* **2013**, *41*, 727-740.
- [125] M. Ko, J. An, H. S. Bandukwala, L. Chavez, T. Aijo, W. A. Pastor, M. F. Segal, H. Li, K. P. Koh, H. Lahdesmaki, P. G. Hogan, L. Aravind, A. Rao, *Nature* **2013**, *497*, 122-126.
- [126] V. K. Ponnaluri, J. P. Maciejewski, M. Mukherji, *Biochem. Biophys. Res. Commun.* **2013**, *436*, 115-120.
- [127] K. Williams, J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. Cloos, J. Rappsilber, K. Helin, *Nature* **2011**, *473*, 343-348.
- [128] H. Wu, Y. Zhang, *Cell Cycle* **2011**, *10*, 2428-2436.
- [129] Q. Chen, Y. Chen, C. Bian, R. Fujiki, X. Yu, *Nature* **2013**, *493*, 561-564.
- [130] F. T. Shi, H. Kim, W. Lu, Q. He, D. Liu, M. A. Goodell, M. Wan, Z. Songyang, *J. Biol. Chem.* **2013**, *288*, 20776-20784.
- [131] R. Kadamb, S. Mittal, N. Bansal, H. Batra, D. Saluja, *Eur. J. Cell Biol.* **2013**, *92*, 237-246.
- [132] J. Ahringer, *Trends Genet.* **2000**, *16*, 351-356.
- [133] R. A. Silverstein, K. Ekwall, *Curr. Genet.* **2005**, *47*, 1-17.
- [134] A. Grzenda, G. Lomberk, J. S. Zhang, R. Urrutia, *Biochim. Biophys. Acta* **2009**, *1789*, 443-450.
- [135] O. Yildirim, R. Li, J.-H. Hung, P. B. Chen, X. Dong, L.-S. Ee, Z. Weng, O. J. Rando, T. G. Fazzio, *Cell* **2011**, *147*, 1498-1510.
- [136] J. Ramirez, J. Hagman, *Epigenetics* **2009**, *4*, 532-536.
- [137] K. Meier, A. Brehm, *Epigenetics* **2014**, *9*, 1485-1495.
- [138] M. P. Torchy, A. Hamiche, B. P. Klaholz, *Cell. Mol. Life. Sci.* **2015**, *72*, 2491-2507.
- [139] B. A. Lewis, J. A. Hanover, *J. Biol. Chem.* **2014**, *289*, 34440-34448.

-
- [140] G. W. Hart, C. Slawson, G. Ramirez-Correa, O. Lagerlof, *Annu. Rev. Biochem.* **2011**, *80*, 825-858.
 - [141] R. Deplus, B. Delatte, M. K. Schwinn, M. Defrance, J. Mendez, N. Murphy, M. A. Dawson, M. Volkmar, P. Putmans, E. Calonne, A. H. Shih, R. L. Levine, O. Bernard, T. Mercher, E. Solary, M. Urh, D. L. Daniels, F. Fuks, *EMBO J.* **2013**, *32*, 645-655.
 - [142] P. Vella, A. Scelfo, S. Jammula, F. Chiacchiera, K. Williams, A. Cuomo, A. Roberto, J. Christensen, T. Bonaldi, K. Helin, D. Pasini, *Mol. Cell* **2013**, *49*, 645-656.
 - [143] C. Bauer, K. Gobel, N. Nagaraj, C. Colantuoni, M. Wang, U. Müller, E. Kremmer, A. Rottach, H. Leonhardt, *J. Biol. Chem.* **2015**, *290*, 4801-4812.
 - [144] A. Balasubramani, A. Rao, *Mol. Cell* **2013**, *49*, 618-619.
 - [145] Q. Zhang, X. Liu, W. Gao, P. Li, J. Hou, J. Li, J. Wong, *J. Biol. Chem.* **2014**, *289*, 5986-5996.
 - [146] R. Fujiki, W. Hashiba, H. Sekine, A. Yokoyama, T. Chikanishi, S. Ito, Y. Imai, J. Kim, H. H. He, K. Igarashi, J. Kanno, F. Ohtake, H. Kitagawa, R. G. Roeder, M. Brown, S. Kato, *Nature* **2011**, *480*, 557-560.
 - [147] T. P. Gu, F. Guo, H. Yang, H. P. Wu, G. F. Xu, W. Liu, Z. G. Xie, L. Shi, X. He, S. G. Jin, K. Iqbal, Y. G. Shi, Z. Deng, P. E. Szabo, G. P. Pfeifer, J. Li, G. L. Xu, *Nature* **2011**, *477*, 606-610.
 - [148] K. Iqbal, S.-G. Jin, G. P. Pfeifer, P. E. Szabo, *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 3642-3647.
 - [149] A. Inoue, L. Shen, Q. Dai, C. He, Y. Zhang, *Cell Res.* **2011**, *21*, 1670-1676.
 - [150] W. Mayer, A. Niveleau, J. Walter, R. Fundele, T. Haaf, *Nature* **2000**, *403*, 501-502.
 - [151] C. Li, Y. Lan, L. Schwartz-Orbach, E. Korol, M. Tahiliani, T. Evans, M. G. Goll, *Cell Rep.* **2015**, *12*, 1133-1143.
 - [152] Z. D. Smith, M. M. Chan, T. S. Mikkelsen, H. Gu, A. Gnirke, A. Regev, A. Meissner, *Nature* **2012**, *484*, 339-344.
 - [153] X. S. Ma, X. G. Wang, L. Qin, C. L. Song, F. Lin, J. M. Song, C. C. Zhu, H. L. Liu, *Genet. Mol. Res.* **2014**, *13*, 8632-8639.
 - [154] S. Guibert, T. Forne, M. Weber, *Genome Res.* **2012**, *22*, 633-641.
 - [155] S. Seisenberger, S. Andrews, F. Krueger, J. Arand, J. Walter, F. Santos, C. Popp, B. Thienpont, W. Dean, W. Reik, *Mol. Cell* **2012**, *48*, 849-862.
 - [156] H. G. Leitch, W. W. Tang, M. A. Surani, *Curr. Top. Dev. Biol.* **2013**, *104*, 149-187.
 - [157] S. Gkoutela, K. X. Zhang, T. A. Shafiq, W. W. Liao, J. Hargan-Calvopina, P. Y. Chen, A. T. Clark, *Cell* **2015**, *161*, 1425-1436.
 - [158] I. B. Zovkic, M. C. Guzman-Karlsson, J. D. Sweatt, *Learn. Mem.* **2013**, *20*, 61-74.
 - [159] M. Santiago, C. Antunes, M. Guedes, N. Sousa, C. J. Marques, *Genomics* **2014**, *104*, 334-340.
 - [160] L. Wen, F. Tang, *Genomics* **2014**, *104*, 341-346.
 - [161] W. Sun, L. Zang, Q. Shu, X. Li, *Genomics* **2014**, *104*, 347-351.
 - [162] J. Shin, G. L. Ming, H. Song, *Phil. Trans. R. Soc. B* **2014**, *369*, doi: 10.1098/rstb.2013.0512.
 - [163] T. Kato, K. Iwamoto, *Neuropharmacology* **2014**, *80*, 133-139.
 - [164] M. M. Dawlaty, A. Breiling, T. Le, G. Raddatz, M. I. Barrasa, A. W. Cheng, Q. Gao, B. E. Powell, Z. Li, M. Xu, K. F. Faull, F. Lyko, R. Jaenisch, *Dev. Cell* **2013**, *24*, 310-323.
 - [165] H. Yang, H. Lin, H. Xu, L. Zhang, L. Cheng, B. Wen, J. Shou, K. Guan, Y. Xiong, D. Ye, *Cell Res.* **2014**, *24*, 1017-1020.
 - [166] K. Blaschke, K. T. Ebata, M. M. Karimi, J. A. Zepeda-Martinez, P. Goyal, S. Mahapatra, A. Tam, D. J. Laird, M. Hirst, A. Rao, M. C. Lorincz, M. Ramalho-Santos, *Nature* **2013**, *500*, 222-226.
 - [167] R. Yin, S.-Q. Mao, B. Zhao, Z. Chong, Y. Yang, C. Zhao, D. Zhang, H. Huang, J. Gao, Z. Li, Y. Jiao, C. Li, S. Liu, D. Wu, W. Gu, Y.-G. Yang, G.-L. Xu, H. Wang, *J. Am. Chem. Soc.* **2013**, *135*, 10396-10403.
 - [168] T. Nakagawa, L. Lv, M. Nakagawa, Y. Yu, C. Yu, A. C. D'Alessio, K. Nakayama, H. Y. Fan, X. Chen, Y. Xiong, *Mol. Cell* **2015**, *57*, 247-260.
 - [169] A. Perera, D. Eisen, M. Wagner, S. K. Laube, A. F. Künzel, S. Koch, J. Steinbacher, E. Schulze, V. Splith, N. Mittermeier, M. Müller, M. Biel, T. Carell, S. Michalakis, *Cell Rep.* **2015**, *11*, 283-294.
 - [170] C. J. Schoenherr, D. J. Anderson, *Curr. Opin. Neurobiol.* **1995**, *5*, 566-571.
 - [171] F. S. Jones, R. Meech, *BioEssays* **1999**, *21*, 372-376.
 - [172] J. M. Coulson, *Curr. Biol.* **2005**, *15*, 665-668.
 - [173] V. V. Lunyak, M. G. Rosenfeld, *Cell* **2005**, *121*, 499-501.
 - [174] I. A. Qureshi, S. Gokhan, M. F. Mehler, *Cell Cycle* **2010**, *9*, 4477-4486.
 - [175] S. E. Goetz, B. Vogelstein, S. R. Hamilton, A. P. Feinberg, *Science* **1985**, *228*, 187-190.
 - [176] M. Kulis, M. Esteller, *Adv. Genet.* **2010**, *70*, 27-56.
 - [177] H. Shen, P. W. Laird, *Cell* **2013**, *153*, 38-55.
 - [178] Y. Bergman, H. Cedar, *Nat. Struct. Mol. Biol.* **2013**, *20*, 274-281.
 - [179] J. S. You, P. A. Jones, *Cancer Cell* **2012**, *22*, 9-20.
 - [180] G. Ficiz, J. G. Gribben, *Genomics* **2014**, *104*, 352-357.

-
- [181] Y. Huang, A. Rao, *Trends Genet.* **2014**, *30*, 464-474.
 - [182] G. P. Pfeifer, W. Xiong, M. A. Hahn, S. G. Jin, *Cell Tissue Res.* **2014**, *356*, 631-641.
 - [183] Y. C. Wu, Z. Q. Ling, *Histol. Histopathol.* **2014**, *29*, 991-997.
 - [184] C. Ye, L. Li, *Cancer Biol. Ther.* **2014**, *15*, 10-15.
 - [185] L. I. Kroeze, B. A. van der Reijden, J. H. Jansen, *Biochim. Biophys. Acta* **2015**, *1855*, 144-154.
 - [186] A. A. Rawluszko-Wieczorek, A. Siera, P. P. Jagodzinski, *Crit. Rev. Oncol. Hematol.* **2015**.
 - [187] L. Scourzic, E. Mouly, O. A. Bernard, *Genome Med.* **2015**, *7*(1):9. doi: 10.1186/s13073-015-0134-6.
 - [188] A. Vasanthakumar, L. A. Godley, *Cancer Genet.* **2015**, *208*, 167-177.
 - [189] C. G. Lian, Y. Xu, C. Ceol, F. Wu, A. Larson, K. Dresser, W. Xu, L. Tan, Y. Hu, Q. Zhan, C. W. Lee, D. Hu, B. Q. Lian, S. Kleffel, Y. Yang, J. Neiswender, A. J. Khorasani, R. Fang, C. Lezcano, L. M. Duncan, R. A. Scolyer, J. F. Thompson, H. Kakavand, Y. Houvras, L. I. Zon, M. C. Mihm, Jr., U. B. Kaiser, T. Schatton, B. A. Woda, G. F. Murphy, Y. G. Shi, *Cell* **2012**, *150*, 1135-1146.
 - [190] C. Liu, L. Liu, X. Chen, J. Shen, J. Shan, Y. Xu, Z. Yang, L. Wu, F. Xia, P. Bie, Y. Cui, X. W. Bian, C. Qian, *PLoS ONE* **2013**, *8*, e62828.
 - [191] C. A. Larsson, G. Cote, A. Quintas-Cardama, *Mol. Cancer Res.* **2013**, *11*, 815-827.
 - [192] M. Ko, Y. Huang, A. M. Jankowska, U. J. Pape, M. Tahiliani, H. S. Bandukwala, J. An, E. D. Lamperti, K. P. Koh, R. Ganetzky, X. S. Liu, L. Aravind, S. Agarwal, J. P. Maciejewski, A. Rao, *Nature* **2010**, *468*, 839-843.
 - [193] Y. Yoshida, *Oncologist* **1996**, *1*, 284-287.
 - [194] C. M. Niemeyer, I. Baumann, *Semin. Hematol.* **2008**, *45*, 60-70.
 - [195] D. Rakheja, S. Konoplev, L. J. Medeiros, W. Chen, *Hum. Pathol.* **2012**, *43*, 1541-1551.
 - [196] W. G. Kaelin, Jr., *Cold Spring Harbor Symp. Quant. Biol.* **2011**, *76*, 335-345.
 - [197] O. Warburg, F. Wind, E. Negelein, *J. Gen. Physiol.* **1927**, *8*, 519-530.
 - [198] K. M. Meldi, M. E. Figueroa, *Pharmacol. Ther.* **2015**, *152*, 42-53.
 - [199] N. R. Rose, M. A. McDonough, O. N. King, A. Kawamura, C. J. Schofield, *Chem. Soc. Rev.* **2011**, *40*, 4364-4397.
 - [200] W. Xu, H. Yang, Y. Liu, Y. Yang, P. Wang, S.-H. Kim, S. Ito, C. Yang, P. Wang, M.-T. Xiao, L.-x. Liu, W.-q. Jiang, J. Liu, J.-y. Zhang, B. Wang, S. Frye, Y. Zhang, Y.-h. Xu, Q.-y. Lei, K.-L. Guan, S.-m. Zhao, Y. Xiong, *Cancer Cell* **2011**, *19*, 17-30.
 - [201] M. E. Figueroa, O. Abdel-Wahab, C. Lu, P. S. Ward, J. Patel, A. Shih, Y. Li, N. Bhagwat, A. Vasanthakumar, H. F. Fernandez, M. S. Tallman, Z. Sun, K. Wolniak, J. K. Peeters, W. Liu, S. E. Choe, V. R. Fantin, E. Paietta, B. Lowenberg, J. D. Licht, L. A. Godley, R. Delwel, P. J. Valk, C. B. Thompson, R. L. Levine, A. Melnick, *Cancer Cell* **2010**, *18*, 553-567.
 - [202] N. Raimundo, B. E. Baysal, G. S. Shadel, *Trends Mol. Med.* **2011**, *17*, 641-649.
 - [203] J. A. McGrath, *J. Investig. Dermatol.* **2003**, *121*, vii.
 - [204] A. Ooi, K. A. Furge, *Chin. J. Cancer.* **2012**, *31*, 413-420.
 - [205] W. M. Linehan, T. A. Rouault, *Clin. Cancer. Res.* **2013**, *19*, 3345-3352.
 - [206] A. King, M. A. Selak, E. Gottlieb, *Oncogene* **2006**, *25*, 4675-4682.
 - [207] C. X. Song, C. Yi, C. He, *Nat. Biotechnol.* **2012**, *30*, 1107-1116.
 - [208] N. Plongthongkum, D. H. Diep, K. Zhang, *Nat. Rev. Genet.* **2014**, *15*, 647-661.
 - [209] C. E. Nestor, J. P. Reddington, M. Benson, R. R. Meehan, *Methods Mol. Biol.* **2014**, *1094*, 243-258.
 - [210] M. Münzel, D. Globisch, C. Trindler, T. Carell, *Org. Lett.* **2010**, *12*, 5671-5673.
 - [211] M. Münzel, U. Lischke, D. Stathis, T. Pfaffeneder, F. A. Gnerlich, C. A. Deiml, S. C. Koch, K. Karaghiosoff, T. Carell, *Chemistry* **2011**, *17*, 13782-13788.
 - [212] A. Szwagierczak, S. Bultmann, C. S. Schmidt, F. Spada, H. Leonhardt, *Nucleic Acids Res* **2010**, *38*, e181.
 - [213] M. J. Booth, E. A. Raiber, S. Balasubramanian, *Chem. Rev.* **2015**, *115*, 2240-2254.
 - [214] J. P. Thomson, J. M. Hunter, C. E. Nestor, D. S. Dunican, R. Terranova, J. G. Moggs, R. R. Meehan, *Nucleic Acids Res.* **2013**, *41*, e206.
 - [215] F. V. Jacinto, E. Ballestar, M. Esteller, *BioTechniques* **2008**, *44*, 35-39.
 - [216] A. L. Sorensen, P. Collas, in *Chromatin Immunoprecipitation Assays: Methods and Protocols*, Vol. 567 (Ed.: P. Collas), **2009**, pp. 249-262.
 - [217] G. Ficiz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews, W. Reik, *Nature* **2011**, *473*, 398-402.
 - [218] H. Wu, A. C. D'Alessio, S. Ito, Z. Wang, K. Cui, K. Zhao, Y. E. Sun, Y. Zhang, *Genes Dev.* **2011**, *25*, 679-684.
 - [219] Y. Xu, F. Wu, L. Tan, L. Kong, L. Xiong, J. Deng, A. J. Barbera, L. Zheng, H. Zhang, S. Huang, J. Min, T. Nicholson, T. Chen, G. Xu, Y. Shi, K. Zhang, Y. G. Shi, *Mol. Cell* **2011**, *42*, 451-464.
 - [220] L. Shen, H. Wu, D. Diep, S. Yamaguchi, A. C. D'Alessio, H.-L. Fung, K. Zhang, Y. Zhang, *Cell* **2013**, *153*, 692-706.

-
- [221] W. A. Pastor, U. J. Pape, Y. Huang, H. R. Henderson, R. Lister, M. Ko, E. M. McLoughlin, Y. Brudno, S. Mahapatra, P. Kapranov, M. Tahiliani, G. Q. Daley, X. S. Liu, J. R. Ecker, P. M. Milos, S. Agarwal, A. Rao, *Nature* **2011**, *473*, 394-397.
- [222] Y. Huang, W. A. Pastor, J. A. Zepeda-Martinez, A. Rao, *Nat. Protoc.* **2012**, *7*, 1897-1908.
- [223] A. B. Robertson, J. A. Dahl, C. B. Vagbo, P. Tripathi, H. E. Krokan, A. Klungland, *Nucleic Acids Res.* **2011**, *39*, e55.
- [224] J. H. Gommersampt, F. Vanleeuwen, A. L. J. Debeer, J. F. G. Vliegenthart, M. Dizdaroglu, J. A. Kowalak, P. F. Crain, P. Borst, *Cell* **1993**, *75*, 1129-1136.
- [225] M. Cross, R. Kieft, R. Sabatini, M. Wilm, M. de Kort, G. A. van der Marel, J. H. van Boom, F. van Leeuwen, P. Borst, *EMBO J.* **1999**, *18*, 6573-6581.
- [226] P. Borst, R. Sabatini, in *Annu. Rev. Microbiol.*, Vol. 62, **2008**, pp. 235-251.
- [227] J. Tomaschewski, H. Gram, J. W. Crabb, W. Ruger, *Nucleic Acids Res.* **1985**, *13*, 7551-7568.
- [228] C.-X. Song, K. E. Szulwach, Y. Fu, Q. Dai, C. Yi, X. Li, Y. Li, C.-H. Chen, W. Zhang, X. Jian, J. Wang, L. Zhang, T. J. Looney, B. Zhang, L. A. Godley, L. M. Hicks, B. T. Lahn, P. Jin, C. He, *Nat. Biotechnol.* **2011**, *29*, 68-72.
- [229] E. A. Raiber, D. Beraldi, G. Ficuz, H. E. Burgess, M. R. Branco, P. Murat, D. Oxley, M. J. Booth, W. Reik, S. Balasubramanian, *Genome Biol.* **2012**, *13*, R69.
- [230] C.-X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S.-Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G.-L. Xu, P. Jin, C. He, *Cell* **2013**, *153*, 678-691.
- [231] R. E. Hardisty, F. Kawasaki, A. B. Sahakyan, S. Balasubramanian, *J. Am. Chem. Soc.* **2015**, *137*, 9270-9272.
- [232] M. Yu, C. X. Song, C. He, *Methods* **2015**, *72*, 16-20.
- [233] M. Frommer, L. E. McDonald, D. S. Millar, C. M. Collis, F. Watt, G. W. Grigg, P. L. Molloy, C. L. Paul, *Proc. Natl. Acad. Sci. U. S. A.* **1992**, *89*, 1827-1831.
- [234] H. Hayatsu, Y. Wataya, K. Kai, S. Iida, *Biochemistry* **1970**, *9*, 2858-2865.
- [235] R. Shapiro, R. E. Servis, M. Welcher, *J. Am. Chem. Soc.* **1970**, *92*, 422-424.
- [236] Y. Huang, W. A. Pastor, Y. Shen, M. Tahiliani, D. R. Liu, A. Rao, *PLoS ONE* **2010**, *5*.
- [237] H. Hayatsu, M. Shiragami, *Biochemistry* **1979**, *18*, 632-637.
- [238] L. Jin, W. Wang, D. Hu, J. Lu, *Phys. Chem. Chem. Phys.* **2014**, *16*, 3573-3585.
- [239] M. J. Booth, M. R. Branco, G. Ficuz, D. Oxley, F. Krueger, W. Reik, S. Balasubramanian, *Science* **2012**, *336*, 934-937.
- [240] M. Yu, G. C. Hon, K. E. Szulwach, C.-X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J.-H. Min, P. Jin, B. Ren, C. He, *Cell* **2012**, *149*, 1368-1380.
- [241] M. J. Booth, G. Marsico, M. Bachman, D. Beraldi, S. Balasubramanian, *Nat. Chem.* **2014**, *6*, 435-440.
- [242] X. Lu, C.-X. Song, K. Szulwach, Z. Wang, P. Weidenbacher, P. Jin, C. He, *J. Am. Chem. Soc.* **2013**, *135*, 9315-9317.
- [243] E. E. Schadt, S. Turner, A. Kasarskis, *Hum. Mol. Genet.* **2010**, *19*, R227-R240.
- [244] J. P. Thomson, A. Fawkes, R. Ottaviano, J. M. Hunter, R. Shukla, H. K. Mjoseng, R. Clark, A. Coutts, L. Murphy, R. R. Meehan, *Sci. Rep.* **2015**, *5*:9778. doi:10.1038/srep09778.
- [245] <http://www.pacificbiosciences.com/products/smrt-technology/>
- [246] B. A. Flusberg, D. R. Webster, J. H. Lee, K. J. Travers, E. C. Olivares, T. A. Clark, J. Korlach, S. W. Turner, *Nat. Methods* **2010**, *7*, 461-465.
- [247] L. Chavez, Y. Huang, K. Luong, S. Agarwal, L. M. Iyer, W. A. Pastor, V. K. Hench, S. A. Frazier-Bowers, E. Korol, S. Liu, M. Tahiliani, Y. Wang, T. A. Clark, J. Korlach, P. J. Pukkila, L. Aravind, A. Rao, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, E5149-5158.
- [248] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. deWinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, S. Turner, *Science* **2009**, *323*, 133-138.
- [249] T. A. Clark, X. Lu, K. Luong, Q. Dai, M. Boitano, S. W. Turner, C. He, J. Korlach, *BMC Biol.* **2013**, *11*, DOI:10.1186/1741-7007-1111-1184.
- [250] C. X. Song, T. A. Clark, X. Y. Lu, A. Kislyuk, Q. Dai, S. W. Turner, C. He, J. Korlach, *Nat. Methods* **2012**, *9*, 75-77.
- [251] T. A. Clark, K. E. Spittle, S. W. Turner, J. Korlach, *Genome integrity* **2011**, *2*, doi:10.1186/2041-9414-1182-1110.

-
- [252] I. D. Vilfan, Y. C. Tsai, T. A. Clark, J. Wegener, Q. Dai, C. Yi, T. Pan, S. W. Turner, J. Korlach, *J. Nanobiotechnol.* **2013**, *11*, doi:10.1186/1477-3155-1111-1188.
- [253] J. J. Kasianowicz, E. Brandin, D. Branton, D. W. Deamer, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 13770-13773.
- [254] B. M. Venkatesan, R. Bashir, *Nat. Nanotechnol.* **2011**, *6*, 615-624.
- [255] D. Branton, D. W. Deamer, A. Marziali, H. Bayley, S. A. Benner, T. Butler, M. Di Ventra, S. Garaj, A. Hibbs, X. Huang, S. B. Jovanovich, P. S. Krstic, S. Lindsay, X. S. Ling, C. H. Mastrangelo, A. Meller, J. S. Oliver, Y. V. Pershin, J. M. Ramsey, R. Riehn, G. V. Soni, V. Tabard-Cossa, M. Wanunu, M. Wiggin, J. A. Schloss, *Nat. Biotechnol.* **2008**, *26*, 1146-1153.
- [256] G. Maglia, A. J. Heron, D. Stoddart, D. Japrun, H. Bayley, *Methods Enzymol.* **2010**, *475*, 591-623.
- [257] M. Wanunu, D. Cohen-Karni, R. R. Johnson, L. Fields, J. Benner, N. Peterman, Y. Zheng, M. L. Klein, M. Drndic, *J. Am. Chem. Soc.* **2011**, *133*, 486-492.
- [258] W.-W. Li, L. Gong, H. Bayley, *Angew. Chem. Int. Ed.* **2013**, *52*, 4350-4355.
- [259] E. V. Wallace, D. Stoddart, A. J. Heron, E. Mikhailova, G. Maglia, T. J. Donohoe, H. Bayley, *Chem. Commun.* **2010**, *46*, 8195-8197.
- [260] A. H. Laszlo, I. M. Derrington, H. Brinkerhoff, K. W. Langford, I. C. Nova, J. M. Samson, J. J. Bartlett, M. Pavlenok, J. H. Gundlach, *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 18904-18909.
- [261] L. Cui, T. H. Chung, D. Tan, X. Sun, X. Y. Jia, *Genomics* **2014**, *104*, 368-375.
- [262] M. T. Zhao, J. J. Whyte, G. M. Hopkins, M. D. Kirk, R. S. Prather, *Cell. Reprogram.* **2014**, *16*, 175-184.
- [263] F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, F. Anselmi, C. Parlato, C. Medana, F. Dal Bello, S. Oliviero, *Cell Rep.* **2015**, doi:10.1016/j.celrep.2015.01.008.
- [264] H. Wu, X. Wu, L. Shen, Y. Zhang, *Nat. Biotechnol.* **2014**, *32*, 1231-1240.
- [265] M. A. Hahn, A. X. Li, X. Wu, G. P. Pfeifer, *Methods Mol. Biol.* **2015**, *1238*, 273-287.
- [266] A. Pettersson, T. H. Chung, D. Tan, X. Sun, X. Y. Jia, *Genome Biol.* **2014**, *15*, doi:10.1186/s13059-13014-10456-13055.
- [267] K. L. Nazor, M. J. Boland, M. Bibikova, B. Klotzle, M. Yu, V. L. Glenn-Pratola, J. P. Schell, R. L. Coleman, M. C. Cabral-da-Silva, U. Schmidt, S. E. Peterson, C. He, J. F. Loring, J. B. Fan, *Genomics* **2014**, *104*, 358-367.
- [268] S.-G. Jin, X. Wu, A. X. Li, G. P. Pfeifer, *Nucleic Acids Res.* **2011**, *39*(12), 5015-5024.
- [269] K. E. Szulwach, X. Li, Y. Li, C.-X. Song, H. Wu, Q. Dai, H. Irier, A. K. Upadhyay, M. Gearing, A. I. Levey, A. Vasanthakumar, L. A. Godley, Q. Chang, X. Cheng, C. He, P. Jin, *Nat. Neurosci.* **2011**, *14*, 1607-1616.
- [270] H. Chen, S. Dzitoyeva, H. Manev, *Restor. Neurol. Neuros.* **2012**, *30*, 237-245.
- [271] L. Chouliaras, D. L. A. van den Hove, G. Kenis, S. Keitel, P. R. Hof, J. van Os, H. W. M. Steinbusch, C. Schmitz, B. P. F. Rutten, *Curr. Alzheimer Res.* **2012**, *9*, 536-544.
- [272] T. Wang, Q. Pan, L. Lin, K. E. Szulwach, C.-X. Song, C. He, H. Wu, S. T. Warren, P. Jin, R. Duan, X. Li, *Hum. Mol. Genet.* **2012**, *21*, 5500-5510.
- [273] M. A. Hahn, R. Qiu, X. Wu, A. X. Li, H. Zhang, J. Wang, J. Jui, S. G. Jin, Y. Jiang, G. P. Pfeifer, Q. Lu, *Cell Rep.* **2013**, *3*, 291-300.
- [274] N. Coppieters, B. V. Dieriks, C. Lill, R. L. Faull, M. A. Curtis, M. Dragunow, *Neurobiol. Aging* **2014**, *35*, 1334-1344.
- [275] L. Wen, X. Li, L. Yan, Y. Tan, R. Li, Y. Zhao, Y. Wang, J. Xie, Y. Zhang, C. Song, M. Yu, X. Liu, P. Zhu, X. Li, Y. Hou, H. Guo, X. Wu, C. He, R. Li, F. Tang, J. Qiao, *Genome Biol.* **2014**, *15*, R49.
- [276] S. F. Field, D. Beraldi, M. Bachman, S. K. Stewart, S. Beck, S. Balasubramanian, *PLoS ONE* **2015**, *10*, e0118202.
- [277] T. F. Kraus, V. Guibourt, H. A. Kretschmar, *J. Neural. Transm.* **2015**, *122*, 1035-1043.
- [278] L. Shen, Y. Zhang, *Curr. Opin. Cell Biol.* **2013**, *25*, 289-296.
- [279] B. Li, M. Carey, J. L. Workman, *Cell* **2007**, *128*, 707-719.
- [280] A. I. Lamond, W. C. Earnshaw, *Science* **1998**, *280*, 547-553.
- [281] H. Stroud, S. Feng, S. M. Kinney, S. Pradhan, S. E. Jacobsen, *Genome Biol.* **2011**, *12*, doi:10.1186/gb-2011-1112-1186-r1154.
- [282] R. Lister, E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao, M. Esteller, C. He, F. G. Haghighi, T. J. Sejnowski, M. M. Behrens, J. R. Ecker, *Science* **2013**, *341*, 1237905.
- [283] K. E. Szulwach, X. Li, Y. Li, C.-X. Song, J. W. Han, S. Kim, S. Namburi, K. Hermetz, J. J. Kim, M. K. Rudd, Y.-S. Yoon, B. Ren, C. He, P. Jin, *PLoS Genet.* **2011**, *7*, e1002154.
- [284] G. A. Maston, S. K. Evans, M. R. Green, in *Annu. Rev. Genomics Hum. Genet.*, Vol. 7, **2006**, pp. 29-59.
- [285] M. Mellen, P. Ayata, S. Dewell, S. Kriaucionis, N. Heintz, *Cell* **2012**, *151*, 1417-1430.

-
- [286] A. Meissner, T. S. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. Zhang, B. E. Bernstein, C. Nusbaum, D. B. Jaffe, A. Gnirke, R. Jaenisch, E. S. Lander, *Nature* **2008**, *454*, 766-770.
 - [287] R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, J. R. Ecker, *Nature* **2009**, *462*, 315-322.
 - [288] C. Coulondre, J. H. Miller, P. J. Farabaugh, W. Gilbert, *Nature* **1978**, *274*, 775-780.
 - [289] A. P. Bird, *Nucleic Acids Res.* **1980**, *8*, 1499-1504.
 - [290] M. A. Hahn, P. E. Szabo, G. P. Pfeifer, *Genomics* **2014**, *104*, 314-323.
 - [291] M. Bulger, M. Groudine, *Cell* **2011**, *144*, 327-339.
 - [292] L. M. Wheldon, A. Abakir, Z. Ferjentsik, T. Dudnakova, S. Strohbuecker, D. Christie, N. Dai, S. Guan, J. M. Foster, I. R. Correa, Jr., M. Loose, J. E. Dixon, V. Sottile, A. D. Johnson, A. Ruzov, *Cell Rep.* **2014**, *7*, 1353-1361.
 - [293] M. W. Kellinger, C.-X. Song, J. Chong, X.-Y. Lu, C. He, D. Wang, *Nat. Struct. Mol. Biol.* **2012**, *19*, 831-833.
 - [294] L. Wang, Y. Zhou, L. Xu, R. Xiao, X. Lu, L. Chen, J. Chong, H. Li, C. He, X. D. Fu, D. Wang, *Nature* **2015**, *523*, 621-625.
 - [295] X. Lu, D. Han, B. S. Zhao, C. X. Song, L. S. Zhang, L. C. Dore, C. He, *Cell Res.* **2015**, *25*, 386-389.
 - [296] M. Iurlaro, G. Ficiz, D. Oxley, E. A. Raiber, M. Bachman, M. J. Booth, S. Andrews, S. Balasubramanian, W. Reik, *Genome Biol.* **2013**, *14*, R119.
 - [297] M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell, S. Balasubramanian, *Nat. Chem. Biol.* **2015**, *11*, 555-557.
 - [298] E. A. Raiber, P. Murat, D. Y. Chirgadze, D. Beraldi, B. F. Luisi, S. Balasubramanian, *Nat. Struct. Mol. Biol.* **2015**, *22*, 44-49.
 - [299] E. Zarakowska, D. Gackowski, M. Foksinski, R. Olinski, *Mutat. Res.-Gen. Tox. En.* **2014**, *764-765*, 58-63.
 - [300] H. J. Lee, T. A. Hore, W. Reik, *Cell Stem Cell* **2014**, *14*, 710-719.
 - [301] F. M. Piccolo, A. G. Fisher, *Trends Cell Biol.* **2014**, *24*, 136-143.
 - [302] P. W. Hill, R. Amouroux, P. Hajkova, *Genomics* **2014**, *104*, 324-333.
 - [303] W. Sun, M. Guan, X. Li, *Stem Cells Dev.* **2014**, *23*, 923-930.
 - [304] S. Yamaguchi, K. Hong, R. Liu, A. Inoue, L. Shen, K. Zhang, Y. Zhang, *Cell Res.* **2013**, *23*, 329-339.
 - [305] S. Kagiwada, K. Kurimoto, T. Hirota, M. Yamaji, M. Saitou, *EMBO J.* **2013**, *32*, 340-353.
 - [306] J. Oswald, S. Engemann, N. Lane, W. Mayer, A. Olek, R. Fundele, W. Dean, W. Reik, J. Walter, *Curr. Biol.* **2000**, *10*, 475-478.
 - [307] M. Wossidlo, T. Nakamura, K. Lepikhov, C. J. Marques, V. Zakhartchenko, M. Boiani, J. Arand, T. Nakano, W. Reik, J. Walter, *Nat. Commun.* **2011**, *2*, doi:10.1038/ncomms1240.
 - [308] A. Inoue, Y. Zhang, *Science* **2011**, *334*, 194-194.
 - [309] H. Hashimoto, Y. Liu, A. K. Upadhyay, Y. Chang, S. B. Howerton, P. M. Vertino, X. Zhang, X. Cheng, *Nucleic Acids Res.* **2012**, *40*, 4841-4849.
 - [310] F. Santos, B. Hendrich, W. Reik, W. Dean, *Dev. Biol.* **2002**, *241*, 172-182.
 - [311] T. Nakamura, Y. Arai, H. Umehara, M. Masuhara, T. Kimura, H. Taniguchi, T. Sekimoto, M. Ikawa, Y. Yoneda, M. Okabe, S. Tanaka, K. Shiota, T. Nakano, *Nat. Cell Biol.* **2007**, *9*, 64-71.
 - [312] T. Nakamura, Y. J. Liu, H. Nakashima, H. Umehara, K. Inoue, S. Matoba, M. Tachibana, A. Ogura, Y. Shinkai, T. Nakano, *Nature* **2012**, *486*, 415-419.
 - [313] H. Kobayashi, T. Sakurai, M. Imai, N. Takahashi, A. Fukuda, O. Yayoi, S. Sato, K. Nakabayashi, K. Hata, Y. Sotomaru, Y. Suzuki, T. Kono, *PLoS Genet.* **2012**, *8*, doi:10.1371/journal.pgen.1002440.
 - [314] P. Hajkova, S. Erhardt, N. Lane, T. Haaf, O. El-Maarri, W. Reik, J. Walter, M. A. Surani, *Mech. Dev.* **2002**, *117*, 15-23.
 - [315] R. Ohno, M. Nakayama, C. Naruse, N. Okashita, O. Takano, M. Tachibana, M. Asano, M. Saitou, Y. Seki, *Development* **2013**, *140*, 2892-2903.
 - [316] Z. D. Smith, A. Meissner, *EMBO J.* **2013**, *32*, 318-321.
 - [317] Y. Kawasaki, J. Lee, A. Matsuzawa, T. Kohda, T. Kaneko-Ishino, F. Ishino, *Sci. Rep.* **2014**, *4*, 3658.
 - [318] Y. Seki, K. Hayashi, K. Itoh, M. Mizugaki, M. Saitou, Y. Matsui, *Dev. Biol.* **2005**, *278*, 440-458.
 - [319] P. Hajkova, K. Ancelin, T. Waldmann, N. Lacoste, U. C. Lange, F. Cesari, C. Lee, G. Almouzni, R. Schneider, M. A. Surani, *Nature* **2008**, *452*, 877-881.
 - [320] R. Margueron, D. Reinberg, *Nature* **2011**, *469*, 343-349.
 - [321] A. A. Mansour, O. Gafni, L. Weinberger, A. Zviran, M. Ayyash, Y. Rais, V. Krupalnik, M. Zerbib, D. Amann-Zalcenstein, I. Maza, S. Geula, S. Viukov, L. Holtzman, A. Pribluda, E. Canaani, S. Horn-Saban, I. Amit, N. Novershtern, J. H. Hanna, *Nature* **2012**, *488*, 409-413.
 - [322] M. Bostick, J. K. Kim, P.-O. Esteve, A. Clark, S. Pradhan, S. E. Jacobsen, *Science* **2007**, *317*, 1760-1764.
 - [323] X. Liu, Q. Gao, P. Li, Q. Zhao, J. Zhang, J. Li, H. Koseki, J. Wong, *Nat. Commun.* **2013**, *4*, 1563.

-
- [324] S. Yamaguchi, K. Hong, R. Liu, L. Shen, A. Inoue, D. Diep, K. Zhang, Y. Zhang, *Nature* **2012**, 492, 443-447.
 - [325] A. Bird, *Nat. Immunol.* **2003**, 4, 208-209.
 - [326] D. Bruniquel, R. H. Schwartz, *Nat. Immunol.* **2003**, 4, 235-240.
 - [327] K. Martinowich, D. Hattori, H. Wu, S. Fouse, F. He, Y. Hu, G. P. Fan, Y. E. Sun, *Science* **2003**, 302, 890-893.
 - [328] S. Kangaspeska, B. Stride, R. Metivier, M. Polycarpou-Schwarz, D. Ibberson, R. P. Carmouche, V. Benes, F. Gannon, G. Reid, *Nature* **2008**, 452, 112-115.
 - [329] F. D. Lubin, T. L. Roth, J. D. Sweatt, *J. Neurosci.* **2008**, 28, 10576-10586.
 - [330] R. Metivier, R. Gallais, C. Tiffocche, C. Le Peron, R. Z. Jurkowska, R. P. Carmouche, D. Ibberson, P. Barath, F. Demay, G. Reid, V. Benes, A. Jeltsch, F. Gannon, G. Salbert, *Nature* **2008**, 452, 45-50.
 - [331] S. K. Ooi, T. H. Bestor, *Cell* **2008**, 133, 1145-1148.
 - [332] D. P. Gavin, K. A. Chase, R. P. Sharma, *Neuropharmacology* **2013**, 75, 233-245.
 - [333] N. N. Karpova, *Neuropharmacology* **2014**, 76, 709-718.
 - [334] I. A. Qureshi, M. F. Mehler, *Phil. Trans. R. Soc. B* **2014**, 369, doi:10.1098/rstb.2013.0506.
 - [335] P. Bekinschtein, M. Cammarota, L. M. Igaz, L. R. M. Bevilacqua, I. Izquierdo, J. H. Medina, *Neuron* **2007**, 53, 261-277.
 - [336] D. K. Ma, M.-H. Jang, J. U. Guo, Y. Kitabatake, M.-I. Chang, N. Pow-Anpongkul, R. A. Flavell, B. Lu, G.-I. Ming, H. Song, *Science* **2009**, 323, 1074-1077.
 - [337] C. A. Miller, J. D. Sweatt, *Neuron* **2007**, 53, 857-869.
 - [338] J. U. Guo, D. K. Ma, H. Mo, M. P. Ball, M. H. Jang, M. A. Bonaguidi, J. A. Balazer, H. L. Eaves, B. Xie, E. Ford, K. Zhang, G. L. Ming, Y. Gao, H. Song, *Nat. Neurosci.* **2011**, 14, 1345-1351.
 - [339] T. L. Roth, J. D. Sweatt, *Curr. Opin. Neurobiol.* **2009**, 19, 336-342.
 - [340] D. Schubeler, *Nature* **2015**, 517, 321-326.
 - [341] E. Li, C. Beard, R. Jaenisch, *Nature* **1993**, 366, 362-365.
 - [342] W. Reik, W. Dean, J. Walter, *Science* **2001**, 293, 1089-1093.
 - [343] S. K. T. Ooi, T. H. Bestor, *Cell* **2008**, 133, 1145-1148.
 - [344] L. Shen, C.-X. Song, C. He, Y. Zhang, *Annu. Rev. Biochem.* **2014**, 83, 585-614.
 - [345] G. L. Xu, C. P. Walsh, *BMB Rep.* **2014**, 47, 609-618.
 - [346] A. Breiling, F. Lyko, *Epigenetics Chromatin* **2015**, 8, 24.
 - [347] G. L. Dianov, U. Hubscher, *Nucleic Acids Res.* **2013**, 41, 3483-3490.
 - [348] H. E. Krokan, M. Bjoras, *Cold Spring Harb. Perspect. Biol.* **2013**, 5, a012583.
 - [349] M. D. Wyatt, *Adv. Cancer Res.* **2013**, 119, 63-106.
 - [350] S. R. Dalton, A. Bellacosa, *Epigenomics-UK* **2012**, 4, 459-467.
 - [351] L. H. Pearl, *Mutation Research-DNA Repair* **2000**, 460, 165-181.
 - [352] D. Cortazar, C. Kunz, Y. Saito, R. Steinacher, P. Schär, *DNA Repair* **2007**, 6, 489-504.
 - [353] P. Gallinari, J. Jiricny, *Nature* **1996**, 383, 735-738.
 - [354] S. Um, M. Harbers, A. Benecke, B. Pierrat, R. Losson, P. Chambon, *J. Biol. Chem.* **1998**, 273, 20728-20736.
 - [355] D. S. Chen, M. J. Lucey, F. Phoenix, J. Lopez-Garcia, S. M. Hart, R. Losson, L. Buluwela, R. C. Coombes, P. Chambon, P. Schär, S. Ali, *J. Biol. Chem.* **2003**, 278, 38586-38592.
 - [356] Y. Q. Li, P. Z. Zhou, X. D. Zheng, C. P. Walsh, G. L. Xu, *Nucleic Acids Res.* **2007**, 35, 390-400.
 - [357] M. Tini, A. Benecke, R. M. Evans, P. Chambon, *Mol. Cell* **2002**, 9, 265-277.
 - [358] D. Cortazar, C. Kunz, J. Selfridge, T. Lettieri, Y. Saito, E. MacDougall, A. Wirz, D. Schürmann, A. L. Jacobs, F. Siegrist, R. Steinacher, J. Jiricny, A. Bird, P. Schär, *Nature* **2011**, 470, 419-423.
 - [359] B. Zhu, Y. Zheng, D. Hess, H. Angliker, S. Schwarz, M. Siegmann, S. Thiry, J. P. Jost, *Proc. Natl. Acad. Sci. U. S. A.* **2000**, 97, 5135-5139.
 - [360] H. Zhang, J. K. Zhu, *Cold Spring Harbor Symp. Quant. Biol.* **2012**, 77, 161-173.
 - [361] L. Zhang, X. Lu, J. Lu, H. Liang, Q. Dai, G.-L. Xu, C. Luo, H. Jiang, C. He, *Nat. Chem. Biol.* **2012**, 8, 328-330.
 - [362] A. B. Sjolund, A. G. Senejani, J. B. Sweasy, *Mutat. Res. Fundam. Mol. Mech. Mutagen.* **2013**, 743, 12-25.
 - [363] H. D. Morgan, W. Dean, H. A. Coker, W. Reik, S. K. Petersen-Mahrt, *J. Biol. Chem.* **2004**, 279, 52353-52360.
 - [364] N. Navaratnam, R. Sarwar, *Int. J. Hematol.* **2006**, 83, 195-200.
 - [365] N. Bhutani, J. J. Brady, M. Damian, A. Sacco, S. Y. Corbel, H. M. Blau, *Nature* **2010**, 463, 1042-1047.
 - [366] S. G. Conticello, *Genome Biol.* **2008**, 9, doi:10.1186/gb-2008-1189-1186-1229.
 - [367] C. Prochnow, R. Bransteitter, X. S. Chen, *Sci. China C Life Sci.* **2009**, 52, 893-902.
 - [368] D. M. Franchini, S. K. Petersen-Mahrt, *Epigenomics-UK* **2014**, 6, 427-443.

-
- [369] R. S. Harris, M. T. Liddament, *Nat. Rev. Immunol.* **2004**, *4*, 868-877.
 - [370] A. Durandy, *Eur. J. Immunol.* **2003**, *33*, 2069-2073.
 - [371] C. E. Hamilton, F. N. Papavasiliou, B. R. Rosenberg, *RNA Biol.* **2010**, *7*, 220-228.
 - [372] M. Cogne, *Biomed. J.* **2013**, *36*, 259-268.
 - [373] K. Rai, I. J. Huggins, S. R. James, A. R. Karpf, D. A. Jones, B. R. Cairns, *Cell* **2008**, *135*, 1201-1212.
 - [374] C. Popp, W. Dean, S. Feng, S. J. Cokus, S. Andrews, M. Pellegrini, S. E. Jacobsen, W. Reik, *Nature* **2010**, *463*, 1101-1105.
 - [375] P. Revy, T. Muto, Y. Levy, F. Geissmann, A. Plebani, O. Sanal, N. Catalan, M. Forveille, R. Dufourcq-Labelouse, A. Gennery, I. Tezcan, F. Ersoy, H. Kayserili, A. G. Ugazio, N. Brousse, M. Muramatsu, L. D. Notarangelo, K. Kinoshita, T. Honjo, A. Fischer, A. Durandy, *Cell* **2000**, *102*, 565-575.
 - [376] C. S. Nabel, H. Jia, Y. Ye, L. Shen, H. L. Goldschmidt, J. T. Stivers, Y. Zhang, R. M. Kohli, *Nat. Chem. Biol.* **2012**, *8*, 751-758.
 - [377] G. Rangam, K.-M. Schmitz, A. J. A. Cobb, S. K. Petersen-Mahrt, *PLoS ONE* **2012**, *7*, doi:10.1371/journal.pone.0043279.
 - [378] M. A. Grillo, S. Colombatto, *Amino Acids* **2005**, *28*, 357-362.
 - [379] M. R. Bauerle, E. L. Schwalm, S. J. Booker, *J. Biol. Chem.* **2015**, *290*, 3995-4002.
 - [380] S. Schiesser, T. Pfaffeneder, K. Sadeghian, B. Hackner, B. Steigenberger, A. S. Schröder, J. Steinbacher, G. Kashiwazaki, G. Höfner, K. T. Wanner, C. Ochsenfeld, T. Carell, *J. Am. Chem. Soc.* **2013**, *135*, 14593-14599.
 - [381] C. K. Liu, C. A. Hsu, M. T. Abbott, *Arch. Biochem. Biophys.* **1973**, *159*, 180-187.
 - [382] J. A. Smiley, M. Kundracik, D. A. Landfried, V. R. Barnes, A. A. Axhemi, *Biochim. Biophys. Acta* **2005**, *1723*, 256-264.
 - [383] S. Schiesser, B. Hackner, T. Pfaffeneder, M. Müller, C. Hagemeier, M. Truss, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 6516-6520.
 - [384] M. Münzel, D. Globisch, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 6460-6468.
 - [385] M. R. Branco, G. Ficiz, W. Reik, *Nat. Rev. Genet.* **2012**, *13*, 7-13.
 - [386] Y. Fu, C. He, *Curr. Opin. Chem. Biol.* **2012**, *16*, 516-524.
 - [387] D. Schubeler, *Nature* **2015**, *517*, 321-326.
 - [388] B. Hendrich, A. Bird, *Mol. Cell. Biol.* **1998**, *18*, 6538-6547.
 - [389] S. A. Sarraf, I. Stancheva, *Mol. Cell* **2004**, *15*, 595-605.
 - [390] H. H. Ng, Y. Zhang, B. Hendrich, C. A. Johnson, B. M. Turner, H. Erdjument-Bromage, P. Tempst, D. Reinberg, A. Bird, *Nat. Genet.* **1999**, *23*, 58-61.
 - [391] X. S. Nan, H. H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman, A. Bird, *Nature* **1998**, *393*, 386-389.
 - [392] H. H. Ng, P. Jeppesen, A. Bird, *Mol. Cell. Biol.* **2000**, *20*, 1394-1406.
 - [393] P. L. Jones, G. J. C. Veenstra, P. A. Wade, D. Vermaak, S. U. Kass, N. Landsberger, J. Strouboulis, A. P. Wolffe, *Nat. Genet.* **1998**, *19*, 187-191.
 - [394] H. H. Ng, A. Bird, *Trends Biochem. Sci.* **2000**, *25*, 121-126.
 - [395] J. C. Rice, C. D. Allis, *Curr. Opin. Cell Biol.* **2001**, *13*, 263-273.
 - [396] S.-G. Jin, S. Kadam, G. P. Pfeifer, *Nucleic Acids Res.* **2010**, *38*, e125.
 - [397] H. Hashimoto, Y. Liu, A. K. Upadhyay, Y. Chang, S. B. Howerton, P. M. Vertino, X. Zhang, X. Cheng, *Nucleic Acids Res.* **2012**, *40*, 4841-4849.
 - [398] V. Valinluck, H. H. Tsai, D. K. Rogstad, A. Burdzy, A. Bird, L. C. Sowers, *Nucleic Acids Res.* **2004**, *32*, 4100-4108.
 - [399] N. P. Blackledge, J. P. Thomson, P. J. Skene, *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a018648.
 - [400] H. Hashimoto, P. M. Vertino, X. Cheng, *Epigenomics-UK* **2010**, *2*, 657-669.
 - [401] H. K. Long, N. P. Blackledge, R. J. Klose, *Biochem. Soc. Trans.* **2013**, *41*, 727-740.
 - [402] N. P. Blackledge, J. C. Zhou, M. Y. Tolstorukov, A. M. Farcas, P. J. Park, R. J. Klose, *Mol. Cell* **2010**, *38*, 179-190.
 - [403] J. C. Zhou, N. P. Blackledge, A. M. Farcas, R. J. Klose, *Mol. Cell. Biol.* **2012**, *32*, 479-489.
 - [404] C. Xu, C. Bian, R. Lam, A. Dong, J. Min, *Nat. Commun.* **2011**, *2*, doi:10.1038/ncomms1237.
 - [405] L. M. Johnson, M. Bostick, X. Zhang, E. Kraft, I. Henderson, J. Callis, S. E. Jacobsen, *Curr. Biol.* **2007**, *17*, 379-384.
 - [406] C. Frauer, T. Hoffmann, S. Bultmann, V. Casa, M. C. Cardoso, I. Antes, H. Leonhardt, *PLoS ONE* **2011**, *6*, doi:10.1371/journal.pone.0021306.
 - [407] T. Zhou, J. Xiong, M. Wang, N. Yang, J. Wong, B. Zhu, R.-M. Xu, *Mol. Cell* **2014**, *54*, 879-886.
 - [408] M. M. Compton, J. M. Thomson, A. H. Icard, *Apoptosis* **2001**, *6*, 299-314.
 - [409] X. Z. Jiang, H. Toyota, T. Yoshimoto, E. Takada, H. Asakura, J. Mizuguchi, *Apoptosis* **2003**, *8*, 509-519.

-
- [410] H. Toyota, X.-Z. Jiang, H. Asakura, J. Mizuguchi, *Cell. Mol. Biol. Lett.* **2012**, *17*, 36-48.
 - [411] K. Dennis, T. Fan, T. Geiman, Q. S. Yan, K. Muegge, *Genes Dev.* **2001**, *15*, 2940-2944.
 - [412] T. Bartke, M. Vermeulen, B. Xhemalce, S. C. Robson, M. Mann, T. Kouzarides, *Cell* **2010**, *143*, 470-484.
 - [413] H. Santos-Rosa, R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein, N. C. T. Emre, S. L. Schreiber, J. Mellor, T. Kouzarides, *Nature* **2002**, *419*, 407-411.
 - [414] R. Schneider, A. J. Bannister, F. A. Myers, A. W. Thorne, C. Crane-Robinson, T. Kouzarides, *Nat. Cell Biol.* **2004**, *6*, 73-77.
 - [415] G. Pichler, P. Wolf, C. S. Schmidt, D. Meilinger, K. Schneider, C. Frauer, K. Fellingner, A. Rottach, H. Leonhardt, *J. Cell. Biochem.* **2011**, *112*, 2585-2593.
 - [416] M. Dejosez, S. S. Levine, G. M. Frampton, W. A. Whyte, S. A. Stratton, M. C. Barton, P. H. Gunaratne, R. A. Young, T. P. Zwaka, *Genes Dev.* **2010**, *24*, 1479-1484.
 - [417] J. Michaud, V. Praz, N. James Faresse, C. K. Inbaptiste, S. Tyagi, F. Schutz, W. Herr, *Genome Res.* **2013**, *23*, 907-916.
 - [418] F. Uhlmann, J. S. Cai, H. FloresRozas, F. B. Dean, J. Finkelstein, M. Odonnell, J. Hurwitz, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 6521-6526.
 - [419] T. Iwasaki, W. W. Chin, L. Ko, *J. Biol. Chem.* **2001**, *276*, 33375-33383.
 - [420] D. Auboeuf, D. H. Dowhan, X. T. Li, K. Larkin, L. Ko, S. M. Berget, B. W. O'Malley, *Mol. Cell. Biol.* **2004**, *24*, 442-453.
 - [421] M. Takagi, M. J. Absalon, K. G. McLure, M. B. Kastan, *Cell* **2005**, *123*, 49-63.
 - [422] Y. T. Xue, J. M. Wong, G. T. Moreno, M. K. Young, J. Cote, W. D. Wang, *Mol. Cell* **1998**, *2*, 851-861.
 - [423] Y. Zhang, H. H. Ng, H. Erdjument-Bromage, P. Tempst, A. Bird, D. Reinberg, *Genes Dev.* **1999**, *13*, 1924-1935.
 - [424] B. Hendrich, J. Guy, B. Ramsahoye, V. A. Wilson, A. Bird, *Genes Dev.* **2001**, *15*, 710-723.
 - [425] P. Carlsson, M. Mahlapuu, *Dev. Biol.* **2002**, *250*, 1-23.
 - [426] A. S. Lalmansingh, S. Karmakar, Y. Jin, A. K. Nagaich, *BBA-Gene Regul. Mech.* **2012**, *1819*, 707-715.
 - [427] B. A. Benayoun, S. Caburet, R. A. Veitia, *Trends Genet.* **2011**, *27*, 224-232.
 - [428] P. F. Agris, *Nucleic Acids Res* **2004**, *32*, 223-238.
 - [429] P. A. Limbach, P. F. Crain, J. A. McCloskey, *Nucleic Acids Res.* **1994**, *22*, 2183-2196.
 - [430] B. El Yacoubi, M. Bailly, V. de Crecy-Lagard, *Annu. Rev. Genet.* **2012**, *46*, 69-95.
 - [431] Y. Fu, C. He, *Curr. Opin. Chem. Biol.* **2012**, *16*, 516-524.
 - [432] K. Phelps, A. Morris, P. A. Beal, *ACS Chem. Biol.* **2012**, *7*, 100-109.
 - [433] K. Nakano, T. Nakao, K. H. Schram, W. M. Hammargren, T. D. McClure, M. Katz, E. Petersen, *Clin. Chim. Acta* **1993**, *218*, 169-183.
 - [434] J. Ofengand, M. Del Campo, Y. Kaya, *Methods* **2001**, *25*, 365-373.
 - [435] D. Globisch, PhD thesis, Ludwig-Maximilians-Universität (München, Germany), **2011**.
 - [436] D. Globisch, D. Pearson, A. Hienzsche, T. Brückl, M. Wagner, I. Thoma, P. Thumbs, V. Reiter, A. C. Kneuttinger, M. Müller, S. A. Sieber, T. Carell, *Angew. Chem. Int. Ed.* **2011**, *50*, 9739-9742.
 - [437] W. Y. Hsu, W. D. Lin, Y. Tsai, C. T. Lin, H. C. Wang, L. B. Jeng, C. C. Lee, Y. C. Lin, C. C. Lai, F. J. Tsai, *Clin. Chim. Acta* **2011**, *412*, 1861-1866.
 - [438] M. Meusburger, M. Hengesbach, M. Helm, in *RNA and DNA Editing: Methods and Protocols*, Vol. 718 (Ed.: R. Aphasizhev), **2011**, pp. 259-270.
 - [439] S. P. Waghmare, M. J. Dickman, *Anal. Chem.* **2011**, *83*, 4894-4901.
 - [440] C. Brandmayr, M. Wagner, T. Brückl, D. Globisch, D. Pearson, A. C. Kneuttinger, V. Reiter, A. Hienzsche, S. Koch, I. Thoma, P. Thumbs, S. Michalakakis, M. Müller, M. Biel, T. Carell, *Angew. Chem. Int. Ed.* **2012**, *51*, 11162-11165.
 - [441] S. Kellner, A. Ochel, K. Thüning, F. Spenkuch, J. Neumann, S. Sharma, K.-D. Entian, D. Schneider, M. Helm, *Nucleic Acids Res.* **2014**, *42*, e142.
 - [442] H. Gingold, O. Dahan, Y. Pilpel, *Nucleic Acids Res.* **2012**, *40*, 10053-10063.
 - [443] N. Netzer, J. M. Goodenbour, A. David, K. A. Dittmar, R. B. Jones, J. R. Schneider, D. Boone, E. M. Eves, M. R. Rosner, J. S. Gibbs, A. Embry, B. Dolan, S. Das, H. D. Hickman, P. Berglund, J. R. Bennink, J. W. Yewdell, T. Pan, *Nature* **2009**, *462*, 522-526.
 - [444] J. Cox, M. Mann, *Cell* **2007**, *130*, 395-398.
 - [445] B. F. Cravatt, G. M. Simon, J. R. Yates, *Nature* **2007**, *450*, 991-1000.
 - [446] R. Aebersold, M. Mann, *Nature* **2003**, *422*, 198-207.
 - [447] H. C. Eberl, M. Mann, M. Vermeulen, *ChemBioChem* **2011**, *12*, 224-234.
 - [448] D. K. Nomura, M. M. Dix, B. F. Cravatt, *Nat. Rev. Cancer* **2010**, *10*, 630-638.
 - [449] S. Hanash, *Nature* **2003**, *422*, 226-232.
 - [450] M. Buck, M. Connick, B. N. Ames, *Anal. Biochem.* **1983**, *129*, 1-13.

-
- [451] S. Kellner, J. Burhenne, M. Helm, *RNA Biol.* **2010**, *7*, 237-247.
 - [452] C. Wetzel, P. A. Limbach, *J. Proteomics* **2012**, *75*, 3450-3464.
 - [453] A. M. B. Giessing, F. Kirpekar, *J. Proteomics* **2012**, *75*, 3434-3449.
 - [454] B. Thomas, A. V. Akoulitchiev, *Trends Biochem. Sci.* **2006**, *31*, 173-181.
 - [455] F. Klepper, E. M. Jahn, V. Hickmann, T. Carell, *Angew. Chem. Int. Ed.* **2007**, *46*, 2325-2327.
 - [456] G. Dirheimer, W. Baranowski, G. Keith, *Biochimie* **1995**, *77*, 99-103.
 - [457] A. Frickenschmidt, H. Frohlich, D. Bullinger, A. Zell, S. Laufer, C. H. Gleiter, H. Liebich, B. Kammerer, *Biomarkers* **2008**, *13*, 435-449.
 - [458] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri, C. D. Bloomfield, E. S. Lander, *Science* **1999**, *286*, 531-537.
 - [459] U. Scherf, D. T. Ross, M. Waltham, L. H. Smith, J. K. Lee, L. Tanabe, K. W. Kohn, W. C. Reinhold, T. G. Myers, D. T. Andrews, D. A. Scudiero, M. B. Eisen, E. A. Sausville, Y. Pommier, D. Botstein, P. O. Brown, J. N. Weinstein, *Nat. Genet.* **2000**, *24*, 236-244.
 - [460] D. T. Ross, U. Scherf, M. B. Eisen, C. M. Perou, C. Rees, P. Spellman, V. Iyer, S. S. Jeffrey, M. Van de Rijn, M. Waltham, A. Pergamenschikov, J. C. E. Lee, D. Lashkari, D. Shalon, T. G. Myers, J. N. Weinstein, D. Botstein, P. O. Brown, *Nat. Genet.* **2000**, *24*, 227-235.
 - [461] C. M. Perou, T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lonning, A. L. Borresen-Dale, P. O. Brown, D. Botstein, *Nature* **2000**, *406*, 747-752.
 - [462] M. Bittner, P. Meitzer, Y. Chen, Y. Jiang, E. Seftor, M. Hendrix, M. Radmacher, R. Simon, Z. Yakhini, A. Ben-Dor, N. Sampas, E. Dougherty, E. Wang, F. Marincola, C. Gooden, J. Lueders, A. Glatfelter, P. Pollock, J. Carpten, E. Gillanders, D. Leja, K. Dietrich, C. Beaudry, M. Berens, D. Alberts, V. Sondak, N. Hayward, J. Trent, *Nature* **2000**, *406*, 536-540.
 - [463] X. Chen, S. T. Cheung, S. So, S. T. Fan, C. Barry, J. Higgins, K. M. Lai, J. Ji, S. Dudoit, I. O. Ng, M. Van De Rijn, D. Botstein, P. O. Brown, *Mol. Biol. Cell* **2002**, *13*, 1929-1939.
 - [464] M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *Science* **1995**, *270*, 467-470.
 - [465] D. J. Lockhart, E. A. Winzeler, *Nature* **2000**, *405*, 827-836.
 - [466] S. Ramaswamy, P. Tamayo, R. Rifkin, S. Mukherjee, C. H. Yeang, M. Angelo, C. Ladd, M. Reich, E. Latulippe, J. P. Mesirov, T. Poggio, W. Gerald, M. Loda, E. S. Lander, T. R. Golub, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 15149-15154.
 - [467] S. Dudoit, J. Fridlyand, T. P. Speed, *J. Am. Stat. Assoc.* **2002**, *97*, 77-87.
 - [468] D. V. Nguyen, D. M. Rocke, *Bioinformatics* **2002**, *18*, 39-50.
 - [469] L. J. van't Veer, H. Y. Dai, M. J. van de Vijver, Y. D. D. He, A. A. M. Hart, M. Mao, H. L. Peterse, K. van der Kooy, M. J. Marton, A. T. Witteveen, G. J. Schreiber, R. M. Kerkhoven, C. Roberts, P. S. Linsley, R. Bernards, S. H. Friend, *Nature* **2002**, *415*, 530-536.
 - [470] Y. S. Liu, M. P. Patricelli, B. F. Cravatt, *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 14694-14699.
 - [471] N. Jessani, Y. S. Liu, M. Humphrey, B. F. Cravatt, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 10335-10340.
 - [472] N. Jessani, M. Humphrey, W. H. McDonald, S. Niessen, K. Masuda, B. Gangadharan, J. R. Yates, B. M. Mueller, B. F. Cravatt, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 13756-13761.
 - [473] M. B. Eisen, P. T. Spellman, P. O. Brown, D. Botstein, *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 14863-14868.
 - [474] T. Sorlie, C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, T. Thorsen, H. Quist, J. C. Matese, P. O. Brown, D. Botstein, P. E. Lonning, A. L. Borresen-Dale, *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10869-10874.
 - [475] M. C. Alley, D. A. Scudiero, A. Monks, M. L. Hursey, M. J. Czerwinski, D. L. Fine, B. J. Abbott, J. G. Mayo, R. H. Shoemaker, M. R. Boyd, *Cancer Res.* **1988**, *48*, 589-601.
 - [476] A. Monks, D. Scudiero, P. Skehan, R. Shoemaker, K. Paull, D. Vistica, C. Hose, J. Langley, P. Cronise, A. Vaigrowloff, M. Graygoodrich, H. Campbell, J. Mayo, M. Boyd, *J. Natl. Cancer Inst.* **1991**, *83*, 757-766.
 - [477] M. R. Grever, S. A. Schepartz, B. A. Chabner, *Semin. Oncol.* **1992**, *19*, 622-638.
 - [478] S. F. Stinson, M. C. Alley, W. C. Kopp, H. H. Fiebig, L. A. Mullendore, A. F. Pittman, S. Kenney, J. Keller, M. R. Boyd, *Anticancer Res.* **1992**, *12*, 1035-1054.
 - [479] M. R. Boyd, K. D. Pauli, *Drug Dev. Res.* **1995**, *34*, 91-109.
 - [480] M. Lacroix, G. Leclercq, *Breast Cancer Res. Treat.* **2004**, *83*, 249-289.
 - [481] O. N. Ikediobi, H. Davies, G. Bignell, S. Edkins, C. Stevens, S. O'Meara, T. Santarius, T. Avis, S. Barthorpe, L. Brackenbury, G. Buck, A. Butler, J. Clements, J. Cole, E. Dicks, S. Forbes, K. Gray, K. Halliday, R. Harrison, K. Hills, J. Hinton, C. Hunter, A. Jenkinson, D. Jones, V. Kosmidou, R. Lugg, A. Menzies, T. Mironenko, A. Parker, J. Perry, K. Raine, D. Richardson, R. Shepherd, A. Small, R. Smith, H.

- Solomon, P. Stephens, J. Teague, C. Tofts, J. Varian, T. Webb, S. West, S. Widaa, A. Yates, W. Reinhold, J. N. Weinstein, M. R. Stratton, P. A. Futreal, R. Wooster, *Mol. Cancer Ther.* **2006**, *5*, 2606-2612.
- [482] R. H. Shoemaker, *Nat. Rev. Cancer* **2006**, *6*, 813-823.
- [483] S. Nishizuka, L. Charboneau, L. Young, S. Major, W. C. Reinhold, M. Waltham, H. Kourou-Mehr, K. J. Bussey, J. K. Lee, V. Espina, P. J. Munson, E. Petricoin, L. A. Liotta, J. N. Weinstein, *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100*, 14229-14234.
- [484] M. Sprinzl, T. Hartmann, J. Weber, J. Blank, R. Zeidler, *Nucleic Acids Res.* **1989**, *17*, Suppl: r1-172.
- [485] H. Grosjean, M. Sprinzl, S. Steinberg, *Biochimie* **1995**, *77*, 139-141.
- [486] P. F. Crain, J. A. McCloskey, *Nucleic Acids Res.* **1996**, *24*, 98-99.
- [487] S. Dunin-Horkawicz, A. Czerwonec, M. J. Gajda, M. Feder, H. Grosjean, J. M. Bujnicki, *Nucleic Acids Res.* **2006**, *34*, D145-D149.
- [488] A. Czerwonec, S. Dunin-Horkawicz, E. Purta, K. H. Kaminska, J. M. Kasprzak, J. M. Bujnicki, H. Grosjean, K. Rother, *Nucleic Acids Res.* **2009**, *37*, D118-121.
- [489] W. A. Cantara, P. F. Crain, J. Rozenski, J. A. McCloskey, K. A. Harris, X. Zhang, F. A. P. Vendeix, D. Fabris, P. F. Agris, *Nucleic Acids Res.* **2011**, *39*, D195-D201.
- [490] A. L. Konevega, *RNA* **2004**, *10*, 90-101.
- [491] G. Kawai, Y. Yamamoto, T. Kamimura, T. Masegi, M. Sekine, T. Hata, T. Iimori, T. Watanabe, T. Miyazawa, S. Yokoyama, *Biochemistry* **1992**, *31*, 1040-1046.
- [492] J. A. Kowalak, J. J. Dalluge, J. A. McCloskey, K. O. Stetter, *Biochemistry* **1994**, *33*, 7869-7876.
- [493] A. Alexandrov, I. Chernyakov, W. Gu, S. L. Hiley, T. R. Hughes, E. J. Grayhack, E. M. Phizicky, *Mol. Cell* **2006**, *21*, 87-96.
- [494] S. Kiesewetter, G. Ott, M. Sprinzl, *Nucleic Acids Res.* **1990**, *18*, 4677-4682.
- [495] J. Putz, C. Florentz, F. Benseler, R. Giege, *Nat. Struct. Biol.* **1994**, *1*, 580-582.
- [496] L. A. Sylvers, K. C. Rogers, M. Shimizu, E. Ohtsuka, D. Soll, *Biochemistry* **1993**, *32*, 3836-3841.
- [497] T. Muramatsu, K. Nishikawa, F. Nemoto, Y. Kuchino, S. Nishimura, T. Miyazawa, S. Yokoyama, *Nature* **1988**, *336*, 179-181.
- [498] E. Madore, C. Florentz, R. Giege, S. Sekine, S. Yokoyama, J. Lapointe, *Eur. J. Biochem.* **1999**, *266*, 1128-1135.
- [499] M. A. Rould, J. J. Perona, T. A. Steitz, *Nature* **1991**, *352*, 213-218.
- [500] J. J. Perona, M. A. Rould, T. A. Steitz, *Biochemistry* **1993**, *32*, 8758-8771.
- [501] A. Schulze, A. L. Harris, *Nature* **2012**, *491*, 364-373.
- [502] G. Zhang, P. Yang, P. Guo, L. Miele, F. H. Sarkar, Z. Wang, Q. Zhou, *Biochim. Biophys. Acta* **2013**, *1836*, 49-59.
- [503] C. Sebastian, *Crit. Rev. Oncog.* **2014**, *19*, 363-382.
- [504] H. Hori, *Front. Genet.* **2014**, *5*, doi:10.3389/fgene.2014.00144.
- [505] P. D'Haeseleer, *Nat. Biotechnol.* **2005**, *23*, 1499-1501.
- [506] K. Pearson, *Philos. Mag.* **1901**, *2*, 559-572.
- [507] A. K. Jain, M. N. Murty, P. J. Flynn, *ACM Comput. Surv.* **1999**, *31*, 264-323.
- [508] R. Xu, D. Wunsch, *IEEE Trans. Neural. Netw.* **2005**, *16*, 645-678.
- [509] G. Yona, W. Dirks, S. Rahman, *Methods Mol. Biol.* **2009**, *541*, 479-509.
- [510] I. Frades, R. Matthiesen, *Methods Mol. Biol.* **2010**, *593*, 81-107.
- [511] M. Eisen, M. de Hoon, Cluster 3.0 Manual, University of Tokyo, Human Genome Center (Tokyo, Japan), 2002; <http://bonsai.hgc.jp/~mdehoon/software/cluster/cluster3.pdf>
- [512] <http://rana.lbl.gov/EisenSoftware.htm>
- [513] <http://rana.lbl.gov/EisenSoftware.htm>
- [514] A. Monks, D. A. Scudiero, G. S. Johnson, K. D. Paull, E. A. Sausville, *Anti-Cancer Drug Des.* **1997**, *12*, 533-541.
- [515] J. Gollub, G. Sherlock, *Methods Enzymol.* **2006**, *411*, 194-213.
- [516] <http://rana.lbl.gov/EisenSoftware.htm>
- [517] <http://rana.lbl.gov/EisenSoftware.htm>
- [518] S. Dzitoyeva, H. Chen, H. Manev, *Neurobiol. Aging* **2012**, *33*, 2881-2891.
- [519] W. Sun, L. Zang, Q. Shu, X. Li, *Genomics* **2014**, *104*, 347-351.
- [520] L. Chouliaras, D. L. A. van den Hove, G. Kenis, S. Keitel, J. van Os, H. W. M. Steinbusch, C. Schmitz, B. P. F. Rutten, *Soc. Neurosc. Abstr. Viewer Itiner. Plan.* **2011**, *41*.
- [521] http://commons.wikimedia.org/wiki/File:Gehirn,_lateral_-_Lobi_deu.svg#mediaviewer/File:Gehirn,_lateral_-_Lobi_deu.svg
- [522] E. Kriukiene, Z. Liutkeviciute, S. Klimasauskas, *Chem. Soc. Rev.* **2012**, *41*, 6916-6930.